Study on Bacterial Protein Synthesis System toward the Incorporation of D-Amino Acid & Synthesis of 2'-deoxy-3'-mercapto-tRNA

Study on Bacterial Protein Synthesis System toward the Incorporation of D-Amino Acid

&

Synthesis of 2'-deoxy-3'-mercapto-tRNA

A dissertation presented

by

Po-Yi Huang

to

Department of Chemistry & Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemistry

Harvard University

Cambridge, Massachusetts

April, 2014

Dissertation advisor: Dr. George Church *Po-Yi Huang*

Study on Bacterial Protein Synthesis System toward the Incorporation of D-Amino Acid

&

Synthesis of 2'-deoxy-3'-mercapto-tRNA

Abstract

Life is anti-entropic and highly organized phenomenon with two characteristics reinforcing each other: homochirality and the stereospecific catalysis of chemical reactions. The exclusive presence of L-amino acids and R-sugars in living world well depict this. Hypothetically, the amino acids and sugars of reverse chirality could form a parallel kingdom which is highly orthogonal to the present world. The components from this mirror kingdom, such as protein or nucleic acid, will be much more resistant to the defensive mechanism of present living system, which could be of great value. Therefore, by gradually rewiring the present bio-machineries, we look to build a bridge leading us to the space of mirror-imaged biomolecules. We begin by investigating protein synthesis with mirror amino acid since most amino acids contain one chiral center to be inversed comparing to sugars. In this work, we analyzed three stages critical for the incorporation of D-amino acid into ribosomal protein synthesis: amino acylation, EF-Tu binding of amino acyl-tRNA and delivery bias, and ribosome catalyzed peptidyl transfer. We have demonstrated that the affinity between EF-Tu and amino acyl-tRNA plays critical role on D-amino acid incorporation, and built a platform aimed to select for ribosome tolerating D-amino acid better.

Non-ribosomal peptide and polyketide synthesis are another important class of modular biomolecules synthesis. Polyketides, which compose one of the largest groups of therapeutic natural products known today, are built by series of modular polyketide synthetases in bacteria, fungi and plants. Intrigued by their interaction with critical pathways in the cell and hence their therapeutic value, researchers have synthesized numerous polyketides either one by one or in a combinatorial manner. However, both these approaches become inefficient when a large set of scaffold-variant libraries of polyketide is required for drug development. In an effort to address this issue, we propose to utilize a bacterial translation system to build polyketide or polyketide-peptide hybrid scaffolds. More specifically, tRNA with a thiol in place of the hydroxyl group at the 3'-terminus is synthesized, loaded with malonyl ketide substrate and subjected into an *in vitro* translation system.

Table of Contents

Acknowledgements

I would like to thank first to my parents, my family and all peoples who have ever offered help to me during the past six years. It has been very delightful to work in Church lab. I would like to thank George for his insightful advisory and my committee members Dr. David Liu, Dr. Greg Verdine and Dr. Jack Szostak as well. The projects are partially sponsored by EMD Merck and Harvard Origin of Life Initiative program, special thanks to Dr. Dimitar Sasselov and Dr. Shou Wang for their support.

Lab members Dr. Luhan Yang, Dr. John Aach, Mr. Alexander Sunguroff, Dr. Harris Wang, Dr. Mike Sismour, Dr. Mike Jewett, Dr. Liangcai Gu, Dr. Jun Li, Dr. Marc Lajoie, Dr. Ben Stranges, Dr. Bobby Dhadwar and Ms. Sara Vassallo all helped me a lot both in experiment and discussion.

A great deal of synthetic work and mass spectrum analysis were done in Cambridge campus. I would like to thanks Dr. Sunia Trauger, Dr. Chris Johnson for their help on mass spectrum analysis. I would like to also thank Dr. Ryan Spoering, Dr. Ramani Ranatunge, Mr. Zach Zinsli and all Chem 100 class members and summer intern members I have worked with: Ms. Fanny Wang, Ms. Marianne Walwema, Ms. Rumbidzai Mushavi, Ms. Sarah Luebke, Mr. Lucas T. Riley, Mr. James Kidd, Ms. Camila Barrios Camacho, Ms. Lina Antounians, Ms. Ally Freedy, Mr. Phil Ngo, Ms. Gillian Farrell, Mr. Kevin Orfield, Ms. Gabriella Paisan, Ms. Chelsea Gilbert, Ms. Katherine Mentzinger, Ms. Jenifer Brown, Ms. Ana Sofia Guerra and Ms. Sarah Farrell, without them I won't be able to carry these projects thus far.

Chapter 1: Introduction

The self-sustention of biological chirality among amino acid and sugar enantiomers

Chirality was first described in 1848 by Louis Pasteur for spontaneously resolved crystals of sodium ammonium tartrate salt having shapes non-superposable to their mirror image (Flack, 2009; Pasteur, 1848). These two chiral tartrate salt crystals preserve their optical activity in solution, in contrast to quartz which loses optical activity when melted. While knowing nothing about molecular structure at that time, Pasteur demonstrated that optical activity can result from intrinsic asymmetry of compounds. Later in 1874, J. H. Van't Hoff and J. A. Le Bel attributed chirality to spatial isomers based on carbon tetrahydron structure hypothesis proposed earlier by Kekulé in 1862 (Hoff and Werner, 1898). This idea widely accepted as it successfully explained many compound's optical activity with known constitutional structure. In the 1800s, several amino acids were isolated from organic source (ex. asparagine from asparagus juice in 1806, cysteine from urinary calculus in 1810, leucine from milk curd in 1819) or differential hydrolysis of protein (ex. glycine from gelatin in 1820), and many of their optical activity were demonstrated (Vickery and Schmidt, 1931). In early 1910s, Emil Fischer resolved several optical isomers of synthetic racemic amino acids (ex. tyrosine, valine, serine and phenylalanine by brucine salt). He along with Franz Hofmeister proved that proteins are polymer of amino acids with amide linkages, by showing close resemblance of synthetic polypeptides to proteins in many properties such as water solubility, reactivity to biuret test, and susceptibility to proteases such as trypsin. Fischer and his colleague also found that racemic peptides (ex. carbethoxyl-glycyl-*dl*-leucine) are hydrolyzed asymmetrically by pancreas extract, and if pure trypsin is used, only peptides with native configuration are digested (Plimmer, 1913). This is probably the earliest evidence that natural machineries such as proteins, not only are composed predominately

by elements with specific configuration, but also interact preferentially with targets containing elements of the same configuration. In addition, nucleosides and the containing sugar are another class of optically active compounds. Not long after the establish of DNA double helix and Watson-Crick base pairing model in 1953, the L-form ribonucleotides were synthesized, followed by L-2'-deoxyoligonucleotides and L-oligonucleotides (Anderson et al., 1984; Holý and Šorm, 1971). Coinciding with Fischer's "Key and Lock" stereospecific hypothesis (Fischer, 1894) for proteins, the L-form DNA doesn't hybridize with complementary oligos of its mirror-form, i.e. D-form.

From Miller's spark discharge experiment (Miller, 1953; Parker et al., 2011), one can speculate that early inorganic production of amino acid were racemic, and presumably other small organic molecules as well. Although it remains elusive how the primitive earth biased toward the current chirality, evidence is found on the spontaneous resolving and concatenating of enantiomeric primordial organic compounds. One example is the polymerization of D- or L-γ-benzylglutamic acid anhydrides. Blout and Idelson found that polymerization elongated from D- (or L)-homopolymers seeds show no preference for D- over L-monomers in a racemic pool. However, when L-homopolymer seed is placed in enriched pools of L-amino acid anhydride, the chain grows faster and longer than when mixed-DL-polymer seed is used. This is due to the stabilization of α-helix conformation in the seed and presumably the intermediate (Idelson and Blout, 1958). Another example is the template directed oligomerization of activated nucleoside. G. F. Joyce found that the presence of poly-D-cytidine-5'-monophosphate template greatly help polymerization of D-guanosine-5'-phosphoimidazole but not the L-form antipode (Joyce et al., 1984) when reactions were carried out separately. However, if the reaction is carried out in racemic mixture, the L-form antipode acts as chain terminator, which blocks the sequential

addition or D-guanosine monophosphate. These effects of enantiomeric cross-inhibitions show that some repetitive structural characteristics formed by homopolymers are critical to their catalytic abilities. This effect explains the necessity of homochirality in primitive forms of life, i.e. macromolecules catalyze their own propagation. In fact, biomolecule homochirality and stereospecific catalysis of its propagation are interdependent and equally important features of living system replication.

How do cells interact with D-amino acids and L-sugars?

Given the potential origin of biological homochirality and its importance to biomacromolecule's self-replication, even it is not clear when and how one chirality kingdom overwhelmed the other, i.e. L-amino acid and D-ribose versus D-amino acid and L-ribose, it is not surprising to see that homochiral biopolymers from one kingdom of chirality evolved defensive mechanisms to prevent the incorporation of units from its enantiomeric antipodes. However, this doesn't exclude the potential that besides the core, self-replicating activity, these homochiral biopolymers can develop beneficiary interaction with certain element of its enantiomeric antipodes. Indeed, some D-amino acids and L-sugar still exist in current organisms of L-amino acid and D-sugar world, and they perform specific functions stereospecifically as well. Example are the use of D-alanine in bacterial cell wall (Neuhaus and Baddiley, 2003), D-serine as a neurotransmitter in brain (Wolosker, 2011) and L-arabinose for glycoprotein in plants (Burget, 2003), and many of them could both be synthesized *de novo* (Heck et al., 1996; Yoshimura and Esak, 2003) and enter catabolism pathway as nitrogen or carbon source (Shimizu et al., 2012; Yurimoto et al., 2000). In monomeric compounds, the role of these enantiomeric antipodes toward present cellular machinery resembles metabolites more than constitutive ones in general.

On the other hand, one would suggest that these enantiomeric antipodes, when appear as polymers or are placed into their antipodal polymers, would be inert toward a large proportion of the present biological machineries, since their interaction based on secondary and tertiary structures would be altered and very likely disrupted. Therefore designed polymers of these enantiomeric antipodes could achieve specific tasks in cells while remained orthogonal – more resistant to degradation, less antigenic and less perturbations to normal cellular functions. Several examples have demonstrated this idea. RNA spiegelmers, which is made out of all L-ribonucleotide, are selected to bind specifically and tightly to protein targets and suppress their function (Vater and Klussmann, 2003). Few spiegelmers targeting diabetes and oncology have entered phase II clinical trials now (ref. NOXXON Pharma website). Several D-amino acid containing peptides of therapeutic potential have also been identified (Sun et al., 2012). In all cases above, they are found to possess much greater nuclease or protease resistant and hence have much longer half-life *in vivo*. It is interesting to note that epimerization of individual amino acids within proteins do not always disrupts their existing interaction with other proteins (Nakagawa et al., 2007; Tugyi et al., 2005), which implies that embedding add-in enantiomeric antipodal module to cellular machineries could selectively change their interaction network.

However, the study of these polymers is greatly limited by their availability. SELEX (Tuerk and Gold, 1990) has been adapted in a genius way to make RNA spiegelmers, but still required to first synthesize the mirror image target chemically, which is very difficult when the target is a large protein. A similar situation is encountered when applying mirror-image phage display technique (Schumacher et al., 1996) to select for D-peptides. We are interested in these properties and potential applications of mirror image biopolymers and therefore seeking to develop a more efficient and broadly applicable way to build them.

Oligonucleotides and oligopeptides, no matter D or L-form, could be chemically synthesized in great lengths and in massive parallel fashion, by either ink-jet printing, electrochemical deprotection techniques or the more recent laser-based electrostatic printing of activated amino acid microparticles (Beyer et al., 2007; Breitling et al., 2009; Gao et al., 2004; Maurer et al., 2006). Despite these advances on synthetic methods could generate oligos containing mirror image residues efficiently, one almost inevitably needs delicate methods to transform these oligos into a functional assembly with higher order structure. These methods are usually carried out by enzymes such as polymerase, ligase, etc., and are mostly stereospecific toward substrates. Although several artificial ribozymes has been created to deal with mirror image substances such as Flexizyme for the amino acylation of tRNA with D-amino acid (Fujino et al., 2013; Goto et al., 2008) and L-hammerhead ribozyme for hydrolysis of target (L)-RNA (Wyszko et al., 2013), the vast majority of enzyme function space of mirror image molecules is still vacant. Following the symmetry, one could imagine these catalysts should exist in mirror space, which has been proved by the observation that chemically synthesized D- and L-HIV1-proteases cut their corresponding D-and L-substrates reciprocally (Milton et al., 1992). This leads us to focus on possible ways of building D-amino acid (D-aa) containing protein.

Ribosomal incorporation of D-amino acids into peptides

The most widely used method for protein production is by cloning the corresponding gene into E. coli and utilize the translation system to express it. In *E. coli* cells, protein is made principally by ribosomal translation of a mRNA template. This translation system is superior to chemical synthesis in several aspects: first, materials required are simple, one only needs a proper strain, a plasmid carrying the gene coding for target protein, and proper media; second, it

is easily programmable and with high fidelity with error rate between 10^{-3}–10^{-4} (Kramer and Farabaugh, 2007); third and most importantly, it can make functional protein under physiological condition in various scales and throughputs both *in vivo* and *in vitro*. In addition, *E. coli.* protein synthesis system has been well studied and documented, one could even artificially reconstitute it from individually prepared principal components (Shimizu et al., 2001). However, as the core function of cell reproduction, it is not surprising to find that along this translation system, mechanisms have evolved to block the incorporation of D-amino acid almost in every move (Ahmad et al., 2013). Quoted from an all-encompassing biochemical study on D-tyrosine's mis-incorporation (Yamane et al., 1981):

> "The maximal selectivity that might be realized would include factors of 25 for aminoacylation (Calendar & Berg, 1966a,b), 25 for ternary complex formation, 10 for EF-Tu•GTP-promoted binding, and 5 for peptidyl transfer, a total discrimination factor greater than 10^4. Despite this formidable barrier to incorporation of D-tyrosine from racemic mixtures, the partial selectivity of the pathway allows the incorporation of D-tyrosine at an appreciable rate if no L-tyrosine is present, as the work of Calendar and Berg has shown"

In brief, besides the action of D-amino acid oxidase and D-aminoacyl-tRNA deacylase, there are three places where D-amino acids are discriminated from L-amino acids in the core translation machinery: aminoacylation of tRNA by aminoacyl-tRNA synthetase (aaRS), formation of ternary complex with EF-Tu-GTP as well as its delivery to ribosome, and the ribosome's acceptance and catalysis of peptide-bond formation. Since the existence of exceptions (Yamane et al., 1981) of the aaRS and EF-Tu recognition, we will address these barriers first before considering ribosome's involvement and discuss our approaches to overcome them. We began with a survey on amino acylation specificity of all 20 aaRS toward D-aa, and then compare the chemical-acylation method and the recent ribozyme-catalyzed acylation method. Next, we tested the overall effect of EF-Tu's binding to D- or L-aa-tRNAs, based on

different tRNA backbones as well as the choice of codon used (Smolskaya et al., 2013). Doi reported that mutations in the amino acid side chain accommodating pocket of EF-Tu could enhance the ternary complex formation and delivery of unnatural aa-tRNA into ribosome (Doi et al., 2007). We applied Doi's strategy to make several EF-Tu mutants with bulky residues mutated to alanine in order to create room for D-aa side chain. There are two aspects of optimizing EF-Tu pocket: the direct one is to improve its binding affinity toward D-aa-tRNA, and the second is to confer D-aa-tRNA protection from hydrolysis (Yamane et al., 1981). The ultimate goal of this part of work is to improve the chance of D-aa-tRNA being delivered to ribosome.

Next, we focused on ribosome engineering. It has long been argued that *E. coli* ribosome can catalyze peptide bond formation with D-amino acid. Early evidence of ribosomal incorporation of D-aa (Yamane et al., 1981) was based on dipeptide formation assay, which has a pitfall that background reactions might contribute significantly to the assay readout. In 1991, Bain et al. demonstrated that in an amber-codon suppressing read-through assay, D-phenylalanine incorporation was not detected compare to glycine (Bain et al., 1991). Later on, similar experiment by Schultz groups showed no incorporation of D-alanine while α,α-dimethyl-amino acid does incorporate at 20% efficiency comparing to L-alanine (Ellman et al., 1992). In 2003, Dedkova et al. showed (Dedkova et al., 2003) above-background signal of D-methionine and D-phenylalanine incorporation and create several mutants with enhanced D-aa incorporation. However, these read-though experiments were performed in cell lysate and no further validation is reported regarding the residue actually incorporated. With the success on reconstitution of protein synthesis system in 2001 (Shimizu et al., 2001), this question was revisited but remained controversial. Tan et al. reported in 2004 that in a tripeptide fMet-(D/L-aa)-Glu synthesis system

with only ribosome, initiation factors, elongation factors and chemical acylated D-aa-tRNA, the signal of tripeptide corresponding to a D-Ala incorporation was turn out proved to be L-Ala by comparing its HPLC trace with standards (Tan et al., 2004). On the other hand, using flexizyme to load all 20 D-aa onto tRNAs, Fujino et al. has shown that in purified translation system (commercialized as PURE system) several D-amino acids such as Ala, Ser, Cys, Met, Thr, His and Phe can easily be incorporated to peptides (Fujino et al., 2013). They used MALDI-TOF mass spectrometry to validate the identity of corresponding incorporated amino acid. However the L-aa may contaminate the D-aa starting material, since mass spectrum doesn't tell the difference between D- and L-configurations. In order to avoid this controversy, we decide to first develop an assay which is not only sensitive enough to detect the amino acid incorporated when at low efficiency but can also distinguish D- or L-configuration. We found the classical amino acid analysis method could be of great value if coupled with the use of HPLC and chiral fluorescence auxiliary reagents. We demonstrate this method by proving the incorporation of D-Ala.

Mutagenesis and selection of ribosome for the ability to incorporate D-amino acid consecutively

Despite several debatable papers reporting single D-amino acid incorporation, incorporation of multiple D-amino acids in a row is never reported. Fujino's data confirmed that wild type ribosome doesn't incorporate D-amino acids consecutively, and a spacer of at least two L-amino acids is required to have the second D-amino acid be taken into the polypeptide chain (Fujino et al., 2013). A single incorporation of D-amino acid in the middle of protein would require two successive reactions, one is its attachment to the previous unit and the other is to the

unit after it. Di- and tripeptide synthesis study in the Cornish lab indicates that although D-aa at A-site could react at comparable yield to that of L-aa, a large fraction of ribosome PTC is stalled when elongating D-aa-tRNA entered P-site (Englander, 2011). Mechanistic study by Rachael Green's group excluded direct participation of ribosome RNA residues in the pepidyl transfer (PT) reaction between peptidyl-tRNA and aa-tRNA, but found that the 2'-OH of A76 of peptidyl-tRNA is critical (Weinger et al., 2004; Youngman et al., 2004). Since the determination of high-resolution crystal structure of ribosome with several aminoacyl and peptidyl-tRNA analogues, several *ab initio* computation works have been done to elucidate the catalytic mechanism in detail. It is now known that the attack of α-amine of aa-tRNA to carbonyl in peptidyl-tRNA forms the rate-limiting transition state (TS), and two water molecules present in close proximity to this substrate transition state play critical roles in the following stage: one mediates proton shuttling from 2'-OH to 3'-O^- leaving group in A76 of peptidyl-tRNA; the other stabilize the negative charge on the ester carbonyl oxygen in TS (Trobro and Åqvist, 2005; Wallin and Åqvist, 2010). It is interesting that in this TS, the amide on the preceding pepidyl unit also form hydrogen bonding with ammonium of aa-tRNA and the second water molecule, which suggests it might also contribute to the stabilization of TS. This view matches the observed lower incorporation efficiency when initiated with N-methyl aa-tRNA (Goto and Suga, 2009), and could also explain why bacterial translation has evolved to initiate with formylated amino acid. However, the presence of proline and the successful initiation with acylated-D-aminoacyl-tRNA in the same paper by Goto et al. suggest that this amide stabilization is probably dispensable. Although mechanistic studies suggests that none of the nucleobases in ribosomal RNA (rRNA) seems to be indispensable either, most previous sequence mutagenesis attempts on this core region, e.g. A2451, U2506, U2585, A2602 or other bases around PT center, yield severely

diminished activities even with single mutation. As we sought to relax the substrate specificity of ribosome to both D- and L-amino acid, we realized that saturated mutagenesis approach would leave vast majority of the library being non-active and mutants closely resemble to wild type sequence always stand out in both control and experimental group. Therefore, in the subsequent attempt we enlarge the region to mutate but lower the mutagenesis rate, hoping that mutants will have more balanced level of activities. We also notice that in several regions around peptide transfer center, rRNA is packed as loop or coil instead of complementary stem, so changing the crowdedness might impact as well as swapping specific inter-base hydrogen bonding network. Taken these ideas together, we create mutated sequences by scripts with predefined rate of insertion, deletion or base change, and exploit the recent technology of massive parallel oligonucleotide synthesis to build mutants. The mutant libraries are expressed *in vivo*, and selected *in vitro* to read-through codons corresponding to D-aa-tRNA supplied. Several interesting surviving mutants after twelve rounds of selections are collected and characterized individually.

Non-ribosomal peptide and polyketide synthesis

Although D-aa is excluded from principle protein translation system, it is a commonly used substrate in non-ribosomal peptide and polyketide synthesis (PKS) pathways. Resembling to peptidyl transfer carried through peptidyl adenosine-3'-O-ester, acyl-transfer in these pathways is carried through a thiol on phosphopantetheine group. The difference is that in ketide synthesis, the nucleophile is an enolate instead of an amine. Although non-ribosomal polyketide /polypeptide synthesis shares similar chemical reactions as ribosomal peptide synthesis, it works in an assembly line fashion in which each enzyme only catalyzes one specific residue's addition

or modification reaction (Ferrer et al., 1999; Gindulyte et al., 2006). This makes it more difficult to program polyketide synthesis system than ribosome system. For example, genetic engineering of the PKS pathway, has a record of building over 50 variants of 6-deoxyerythronolide B (DEBS) (Xue et al., 1999). However, it is only in a few systems like DEBS that we know thorough information about the domain DNA sequences and their spatial orientations in the modules that lead to the consecutive processing, where we could then rationally engineer the pathway (Weissman and Leadlay, 2005). Even with this knowledge, this method is found to be extremely inefficient, since the inter-module interactions and dynamics are still largely unclear. Since PKS has broad substrate category, we wonder if we can transplant the polyketide synthesis unit to ribosomal translation system to collect the merits of both. Here, we report the synthesis of a tRNA with thiol at 3'-terminus, starting with the synthesis of dinucleotide pdCpdA-3-SH, and the preliminary observation on its property toward ribosomal polyketide synthesis.

References

Ahmad, S., Routh, S.B., Kamarthapu, V., Chalissery, J., Muthukumar, S., Hussain, T., Kruparani, S.P., Deshmukh, M. V, and Sankaranarayanan, R. (2013). Mechanism of chiral proofreading during translation of the genetic code. Elife *2*, e01519.

Anderson, D.J., Reischer, R.J., Taylor, A.J., and Wechter, W.J. (1984). Preparation and Characterization of Oligonucleotides of D- and L-2' Deoxyuridine. Nucleosides and Nucleotides *3*, 499–512.

Bain, J.D., Wacker, D.A., Kuo, E.E., and Chamberlin, A.R. (1991). Site-specific incorporation of non-natural residues into peptides: Effect of residue structure on suppression and translation efficiencies. Tetrahedron *47*, 2389–2400.

Beyer, M., Nesterov, A., Block, I., König, K., Felgenhauer, T., Fernandez, S., Leibe, K., Torralba, G., Hausmann, M., Trunk, U., et al. (2007). Combinatorial Synthesis of Peptide Arrays onto a Microchip. Science *318*, 1888.

Breitling, F., Nesterov, A., Stadler, V., Felgenhauer, T., and Bischoff, F.R. (2009). High-density peptide arrays. Mol. Biosyst. *5*, 224–234.

Burget, E.G. (2003). The Biosynthesis of L-Arabinose in Plants: Molecular Cloning and Characterization of a Golgi-Localized UDP-D-Xylose 4-Epimerase Encoded by the MUR4 Gene of Arabidopsis. Plant Cell *15*, 523–531.

Dedkova, L.M., Fahmi, N.E., Golovine, S.Y., and Hecht, S.M. (2003). Enhanced d-Amino Acid Incorporation into Protein by Modified Ribosomes. J. Am. Chem. Soc. *125*, 6616–6617.

Doi, Y., Ohtsuki, T., Shimizu, Y., Ueda, T., and Sisido, M. (2007). Elongation Factor Tu Mutants Expand Amino Acid Tolerance of Protein Biosynthesis System. J. Am. Chem. Soc. *129*, 14458–14462.

Ellman, J.A., Mendel, D., and Schultz, P.G. (1992). Site-specific incorporation of novel backbone structures into proteins. Science *255*, 197–200.

Englander, M.T. (2011). The Ribosome Discriminates the Structure of the Amino Acid at its Peptidyl-Transferase Center. Columbia Univertisty.

Ferrer, J.L., Jez, J.M., Bowman, M.E., Dixon, R.A., and Noel, J.P. (1999). Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. Nat. Struct. Mol. Biol. *6*, 775–784.

Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. Berichte Der Dtsch. Chem. Gesellschaft *27*, 2985–2993.

Flack, H. (2009). Louis Pasteur's discovery of molecular chirality and spontaneous resolution in 1848, together with a complete review of his crystallographic and chemical work. Acta Crystallogr. Sect. A *65*, 371–389.

Fujino, T., Goto, Y., Suga, H., and Murakami, H. (2013). Reevaluation of the d-Amino Acid Compatibility with the Elongation Event in Translation. J. Am. Chem. Soc. *135*, 1830–1837.

Gao, X., Gulari, E., and Zhou, X. (2004). In situ synthesis of oligonucleotide microarrays. Biopolymers *73*, 579–596.

Gindulyte, A., Bashan, A., Agmon, I., Massa, L., Yonath, A., and Karle, J. (2006). The transition state for formation of the peptide bond in the ribosome. Proc. Natl. Acad. Sci. U. S. A. *103*, 13327–13332.

Goto, Y., and Suga, H. (2009). Translation Initiation with Initiator tRNA Charged with Exotic Peptides. J. Am. Chem. Soc. *131*, 5040–5041.

Goto, Y., Ohta, A., Sako, Y., Yamagishi, Y., Murakami, H., and Suga, H. (2008). Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. ACS Chem. Biol. *3*, 120–129.

Heck, S.D., Faraci, W.S., Kelbaugh, P.R., Saccomano, N.A., Thadeio, P.F., and Volkmann, R.A. (1996). Posttranslational amino acid epimerization: enzyme-catalyzed isomerization of amino acid residues in peptide chains. Proc. Natl. Acad. Sci. *93* , 4036–4039.

Hoff, J.H. van't, and Werner, A. (1898). The Arrangement of Atoms in Space (Longmans, Green).

Holý, A., and Šorm, F. (1971). Nucleic acid components and their analogues. CXL. Preparation of 5'-L-ribonucleotides, some of their derivatives, and 2'(3')→5'-homooligo-L-ribonucleotides; coding properties of L-ribonucleoside-containing oligonucleotides. Collect. Czechoslov. Chem. Commun. *36*, 3282–3299.

Idelson, M., and Blout, E.R. (1958). Polypeptides. XVIII.1 A Kinetic Study of the Polymerization of Amino Acid N-Carboxyanhydrides Initiated by Strong Bases. J. Am. Chem. Soc. *80*, 2387–2393.

Joyce, G.F., Visser, G.M., van Boeckel, C.A.A., van Boom, J.H., Orgel, L.E., and van Westrenen, J. (1984). Chiral selection in poly(C)-directed synthesis of oligo(G). Nature *310*, 602–604.

Kramer, E.B., and Farabaugh, P.J. (2007). The frequency of translational misreading errors in E. coli is largely determined by tRNA competition. RNA *13*, 87–96.

Maurer, K., Cooper, J., Caraballo, M., Crye, J., Suciu, D., Ghindilis, A., Leonetti, J.A., Wang, W., Rossi, F.M., Stöver, A.G., et al. (2006). Electrochemically Generated Acid and Its Containment to 100 Micron Reaction Areas for the Production of DNA Microarrays. PLoS One *1*, e34.

Miller, S.L. (1953). A Production of Amino Acids Under Possible Primitive Earth Conditions. Science *117*, 528–529.

Milton, R.C. deL., Milton, S.C.F., and Kent, S.B.H. (1992). Total Chemical Synthesis of a D-Enzyme: The Enantiomers of HIV-1 Protease Show Demonstration of Reciprocal Chiral Substrate Specificity. Science *256*, 1445–1448.

Nakagawa, Y., Kikuchi, H., and Takahashi, H. (2007). Molecular Analysis of TCR and Peptide/MHC Interaction Using P18-I10-Derived Peptides with a Single d-Amino Acid Substitution. Biophys. J. *92*, 2570–2582.

Neuhaus, F.C., and Baddiley, J. (2003). A Continuum of Anionic Charge: Structures and Functions of D-Alanyl-Teichoic Acids in Gram-Positive Bacteria. Microbiol. Mol. Biol. Rev. *67*, 686–723.

Parker, E.T., Cleaves, H.J., Dworkin, J.P., Glavin, D.P., Callahan, M., Aubrey, A., Lazcano, A., and Bada, J.L. (2011). Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment. Proc. Natl. Acad. Sci. U. S. A. *108*, 5526–5531.

Pasteur, L. (1848). Mémoire sur la relation qui peut exister entre la forme cristalline et la composition chimique, et sur la cause de la polarisation rotatoire" (Memoir on the relationship which can exist between crystalline form and chemical composition, and on the cause o. Comptes Rendus l'Académie Des Sci. *26*, 535–538.

Plimmer, R.H.A. (1913). The Chemical Constitution of the Proteins: Synthesis (Longmans, Green & Company).

Schumacher, T.N.M., Mayr, L.M., Jr., D.L.M., Milhollen, M.A., Burgess, M.W., and Kim, P.S. (1996). Identification of D-Peptide Ligands Through Mirror-Image Phage Display. Curr. Med. Chem. *271*, 1854–1857.

Shimizu, T., Takaya, N., and Nakamura, A. (2012). An l-glucose Catabolic Pathway in Paracoccus Species 43P. J. Biol. Chem. *287*, 40448–40456.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001). Cell-free translation reconstituted with purified components. Nat. Biotechnol. *19*, 751–755.

Smolskaya, S., Zhang, Z.J., and Alfonta, L. (2013). Enhanced Yield of Recombinant Proteins with Site-Specifically Incorporated Unnatural Amino Acids Using a Cell-Free Expression System. PLoS One *8*, e68363.

Sun, N., Funke, S.A., and Willbold, D. (2012). Mirror image phage display – Generating stable therapeutically and diagnostically active peptides with biotechnological means. J. Biotechnol. *161*, 121–125.

Tan, Z., Forster, A.C., Blacklow, S.C., and Cornish, V.W. (2004). Amino Acid Backbone Specificity of the Escherichia coli Translation Machinery. J. Am. Chem. Soc. *126*, 12752–12753.

Trobro, S., and Åqvist, J. (2005). Mechanism of peptide bond synthesis on the ribosome. Proc. Natl. Acad. Sci. U. S. A. *102*, 12395–12400.

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505–510.

Tugyi, R., Uray, K., Iván, D., Fellinger, E., Perkins, A., and Hudecz, F. (2005). Partial d-amino acid substitution: Improved enzymatic stability and preserved Ab recognition of a MUC2 epitope peptide. Proc. Natl. Acad. Sci. U. S. A. *102*, 413–418.

Vater, A., and Klussmann, S. (2003). Toward third-generation aptamers: Spiegelmers and their therapeutic prospects. Curr. Opin. Drug Discov. Devel. *6*, 253–261.

Vickery, H.B., and Schmidt, C.L.A. (1931). The History of the Discovery of the Amino Acids. Chem. Rev. *9*, 169–318.

Wallin, G., and Åqvist, J. (2010). The transition state for peptide bond formation reveals the ribosome as a water trap. Proc. Natl. Acad. Sci. *107*, 1888–1893.

Weinger, J.S., Parnell, K.M., Dorner, S., Green, R., and Strobel, S.A. (2004). Substrate-assisted catalysis of peptide bond formation by the ribosome. Nat. Struct. Mol. Biol. *11*, 1101–1106.

Weissman, K.J., and Leadlay, P.F. (2005). Combinatorial biosynthesis of reduced polyketides. Nat. Rev. Microbiol. *3*, 925–936.

Wolosker, H. (2011). Serine racemase and the serine shuttle between neurons and astrocytes. Biochim. Biophys. Acta - Proteins Proteomics *1814*, 1558–1566.

Wyszko, E., Szymański, M., Zeichhardt, H., Müller, F., Barciszewski, J., and Erdmann, V.A. (2013). Spiegelzymes: Sequence Specific Hydrolysis of L-RNA with Mirror Image Hammerhead Ribozymes and DNAzymes. PLoS One *8*, e54741.

Xue, Q., Ashley, G., Hutchinson, C.R., and Santi, D. V (1999). A multiplasmid approach to preparing large libraries of polyketides. Proc. Natl. Acad. Sci. U. S. A. *96*, 11740–11745.

Yamane, T., Miller, D.L., and Hopfield, J.J. (1981). Discrimination between D- and L-tyrosyl transfer ribonucleic acids in peptide chain elongation. Biochemistry *20*, 7059–7064.

Yoshimura, T., and Esak, N. (2003). Amino acid racemases: Functions and mechanisms. J. Biosci. Bioeng. *96*, 103–109.

Youngman, E.M., Brunelle, J.L., Kochaniak, A.B., and Green, R. (2004). The Active Site of the Ribosome Is Composed of Two Layers of Conserved Nucleotides with Distinct Roles in Peptide Bond Formation and Peptide Release. Cell *117*, 589–599.

Yurimoto, H., Hasegawa, T., Sakai, Y., and Kato, N. (2000). Physiological role of the D-amino acid oxidase gene, DAO1, in carbon and nitrogen metabolism in the methylotrophic yeast Candida boidinii. Yeast *16*, 1217–1227.

Chapter 2: From D-amino acid to D-aminoacyl-tRNA and its incorporation to peptide through EF-Tu binding

Summary

We examined three different methods to charge D-amino acid onto tRNA: by aaRS, by flexizyme and by chemical synthesis. Among 19 *E. coli* aaRSs, only three show low but measurable activity toward corresponding D-aa. Comparing to flexizyme system, chemical acylation required more labor-intensive synthetic work, but more handy in daily translation assay once reagents are available. We also examined several approaches to increase D-aa incorporation, effective ones including optimizing the tRNA backbones and relaxing EF-Tu aa-tRNA binding pocket. We used read-through assay and develop chiral HPLC method to validate D-amino acid incorporation.

Introduction

In order to investigate ribosomal incorporation of D-amino acid into peptide, the first step would be to build D-amino acid charged-tRNAs. In 1978, Hecht S. et al. showed first example of chemical acylation by ligating chemically acylated P^1,P^2-bis(5'-adenosyl)diphosphate to tRNAs without 3'-A (tRNA-C-C-$_{3'OH}$) (Hecht et al., 1978). This approach was later modified and popularized by Noren C. et al. in 1989 by the ligation of acylated pCpA dinucleotide (later substituted by pdCpA) to tRNAs without 3'-CA dinucleotide (tRNA-C-$_{3'OH}$). This strategy is applicable to any amino acid. *In vitro* transcribed, unmodified tRNAs are often used for the ligation, due to both the difficulty to specifically remove 3'-dinucleotides in native tRNAs and the lack of efficient method to purify individual tRNA isoacceptor. Application of *in vitro* transcribed tRNA has its potential constrain as well, because it lacks modifications which present on naturally produced tRNA, which will affect its anticodon•codon interaction and likely its binding to EF-Tu or ribosome (Björk, 1996).

Recently, Suga et al. has developed an *in vitro* evolved ribozyme capable of acylating any amino acid to full-length tRNA (Goto et al., 2011). Although the acylation yields vary from 10% to 80% for different amino acid, it allows the use of any tRNA in full-length, without the need to remove nucleotides on its 3'-end.

Our first goal is to incorporate D-amino acid into conventional protein synthesis system, in the presence of aaRS, D-amino acid and all cognate tRNAs. Therefore we first assay the activity of purified aaRS for their ability to charge D-amino acid onto tRNA. This is to evaluate if there is any aaRS could serve as alternative method to load D-aa onto tRNA, and also to prevent, if there is such aaRS, the chance of mis-incorporation of D-aa into wrong position when undesired. In this chapter, we described an assay for aaRS charging based on the catalytic PPi exchange of aaRS upon binding to cognate amino acid and tRNA. We then compare the chemical acylation and the flexizyme acylation approach for the synthesis of D-aa-tRNA. We constructed several *in vitro* transcribed tRNAs with different scaffold sequences and anti-codons to compare their affinity toward EF-Tu as well as their decoding efficiency. We also compared several cell-free protein synthesis conditions in literature to find potential space to improve the delivery of non-aaRS acylated, *in vitro* transcribed tRNAs.

Results

Specificity of aminoacyl-tRNA synthetases

The most direct assay to measure aaRS activity would be to monitor the production of aa-tRNA (Figure 1a). However, this is complicated by its instability to hydrolysis and the limited availability of purified cognate tRNA isoacceptor. In literature, attempt has been made to measure the generation of byproduct AMP (Wu and Hill, 1993), by coupling enzymatic reactions

to transform it to the diminish of NADH absorbance. However, we have found the signal of this assay is barely detectable. Therefore, instead of aa-tRNA, we measure the catalysis of amino acid activation step. It is known that this step is reversible, and aaRS catalyze it in both directions. A simple way is to start reaction with ATP and ^{32}PPi and then measure the production of ^{32}P-ATP over time (Eigner and Loftfield, 1974). In order to avoid the use of hazardous radioactive isotopes, we adapt the strategy developed by Roy et al by using AMPNP, an ATP analogue with non-hydrolysable γ-phosphates, with PPi (Roy, 1983), and detect the generated ATP by luminescence assay (Figure 1b).

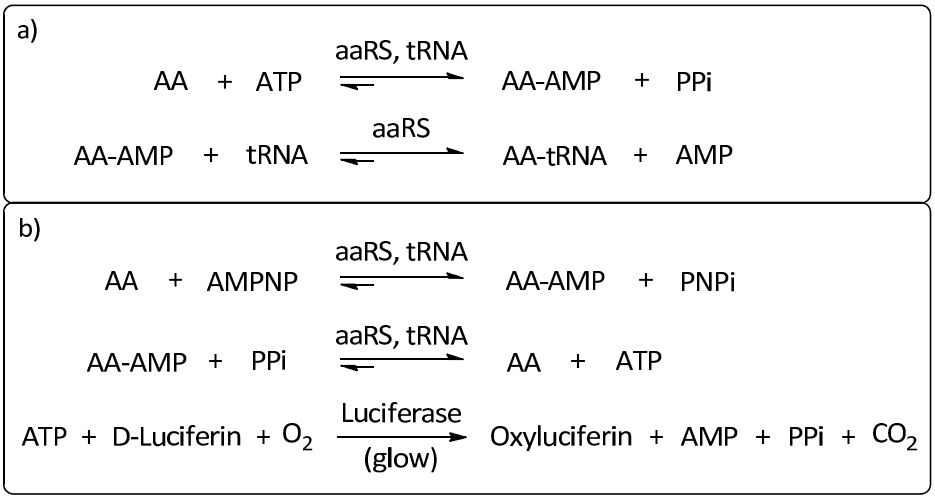


Figure 1.(a) aaRS catalyzed synthesis of tRNA. (b) aaRS activity measured by PPi exchange coupled ATP reading by Luciferase assay.

We set up amino acylation reaction with 10-50 nM of aaRS and 200 μM of AMPNP at 37°C for 30 min, and luminescence is measured right after the solution is mixed with luciferase and D-luciferin. A typical data is shown in Figure 2, the specific activity of aaRS toward D- or L-amino acid is calculated from relative light units (RLUs) by the fitted polynomial curved from ATP standards.

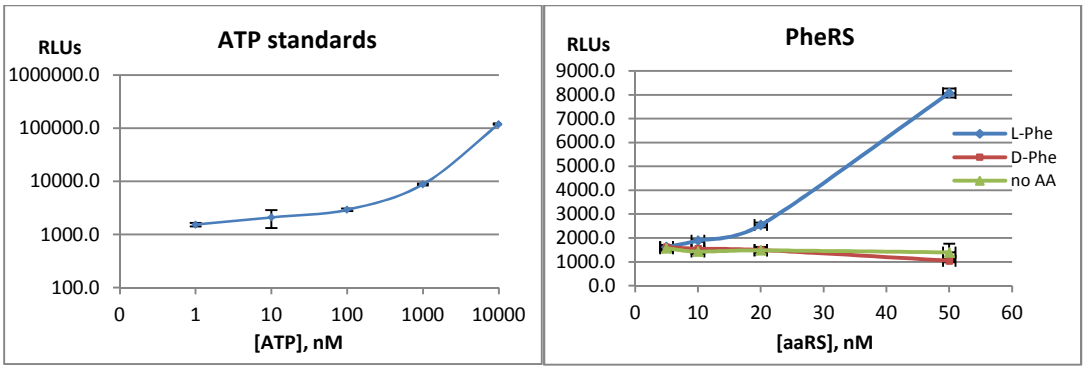


Figure 2. Left, the standard curve of ATP concentration versus the chemical luminescence signal; right, PheRS of various concentrations are tested against D-Phe or L-Phe for ATP generation from PPi and AMPNP.

We applied this assay to all 20 aaRSs, but only 10 out of 20 shows activities when the corresponding L-amino acid presents, they are AlaRS, ThrRS, PheRS, AspRS, AsnRS, HisRS, GlyRS, SerRS and LysRS. When higher concentration of aaRS is used (0.5 to 4.0 μM), CysRS, TrpRS and ArgRS give detectable but low activities. However, since all 20 aaRS enzymes in reconstituted protein synthesis assay work collectively, we suggest the failure of PPi exchange might be the rejection of AMPNP substrate by synthetases. Digging back into literature, we found that some type I aaRS from Baker's yeast indeed can't use AMPNP as substrate (Freist et al., 1980). Therefore, we tried to use another analogue, ATPγS instead. ATPγS gives higher background in luciferase assay, so higher [aaRS] is required to get significant and distinguishable signal from background. Also, we noticed that GluRS and GlnRS both require un-acylated cognate tRNA for the catalysis of PPi exchange (Ravel et al., 1965), so deacylated total tRNA is used in the assay of these two synthetases. The summarized result is shown in Table 1 below.

aaRS	Class	Assay condition	Specific activities (Unit/nmol aaRS)*		
			L-aa (s.d.)	D-aa (s.d.)	No aa (s.d.)
AlaRS	Type II	(a)	1018 (36.6)	n.d.	n.d.
ThrRS	Type II	(a)	29 (1.9)	n.d.	n.d.
PheRS	Type II	(a)	22 (19.0)	n.d.	n.d.
AspRS	Type II	(a)	249 (48.9)	n.d.	n.d.

AsnRS	Type II	(a)	128	(15.6)	n.d.		n.d.	
HisRS	Type II	(a)	115	(29.5)	1.2	(0.7)	2.2	(3.3)
GlyRS	Type II	(a)	142	(64.5)	n.d.		2.8	(3.6)
SerRS	Type II	(a)	70	(21.0)	0.9	(0.6)	1.4	(4.1)
ProRS	Type II	(a)	403	(35.4)	n.d.		n.d.	
LysRS	Type II	(a)	758	(188.7)	7.7	(8.2)	8.4	(2.7)
CysRS	Type I	(b)	7.9	(8.0)	n.d.		n.d.	
TrpRS	Type I	(b)	35.3	(20.4)	20.0	(15.7)	13.0	(2.4)
ArgRS	Type I	(b)	11.8	(3.6)	4.0	(1.7)	0.4	(0.1)
ValRS	Type I	(c)	458.9	(124.4)	n.d.		n.d.	
IleRS	Type I	(c)	911.7	(229.2)	n.d.		n.d.	
LeuRS	Type I	(c)	200.5	(-)	n.d.		n.d.	
MetRS	Type I	(c)	415.3	(144.3)	n.d.		n.d.	
TyrRS	Type I	(c)	1228.3	(506.6)	320.2	(82.8)	n.d.	
GluRS	Type I	(c)	10.3	(2.4)	n.d.		n.d.	
GlnRS	Type I	(c)	1077.1	(246.2)	n.d.		n.d.	

Table 1. Summary of specific activities of aminoacyl-tRNA synthetases promoted by cognate L- or D-amino acids. Assay condition (a) 200 μM AMPNP with 10-50 nM of aaRS, (b) 200 μM AMPNP with 0.5-4.0 μM of aaRS, (c) 200 μM ATPγS with 20-50 nM of aaRS. Specific activities is averaged across the range of [aaRS] tested. *1 unit of aaRS exchanges 1nmol of PPi with AMP-AA to ATP in 30min at 37°C

In agreement with previous report (Soutourina et al., 2000), we find TyrRS, TrpRS and ArgRS have slight to moderate tolerance toward D-form analogue of cognate amino acid. Beside these three synthetases, we didn't detect PPi exchange activity promoted by any other D-amino acids. Our assay shows that the L-form amino acids of these three aaRSs, TyrRS, TrpRS and ArgRS promote catalysis in about 3-4 times faster. However, Soutourina et al. indicates the initial rate of aa-tRNA synthesis with L-form amino acid is about 100 times faster than D-form analogue. The difference suggests that the structural requirement in the second acylation step is more stringent than the first amino acid activation step. And our assay demonstrated that for the vast majority of aaRS, D-amino acids are blocked at the very first stage, so there shouldn't be any concern on toxicity of D-amino acid for cell free protein synthesis application. But on the other hand, using aaRS to charge D-amino acid onto tRNA would be very inefficient if the

corresponding L-aa is present even for D-aa accepting aaRS. In the next section, we will explore two other approaches.

Non-enzymatic acylation and chemical acylation

As we are looking for a general method to efficiently load D-aa onto tRNAs, the Flexizyme system appears to be a good candidate. It was first published in 1996-2000 as an aminoacyl-RNA transferase by Szostak's lab (Lee et al., 2000; Lohse and Szostak, 1996), is later evolved into trans-acting tRNA aminoacylase for cyanomethyl activated L-Phe in 2001 (Saito et al., 2001) and generalized to all amino acids in 2006 by placing the ribozyme's recognition moiety on carboxyl-terminus activating group (Murakami et al., 2006). The general scheme of this amino acid activation and transacylation system is shown in Figure 3.

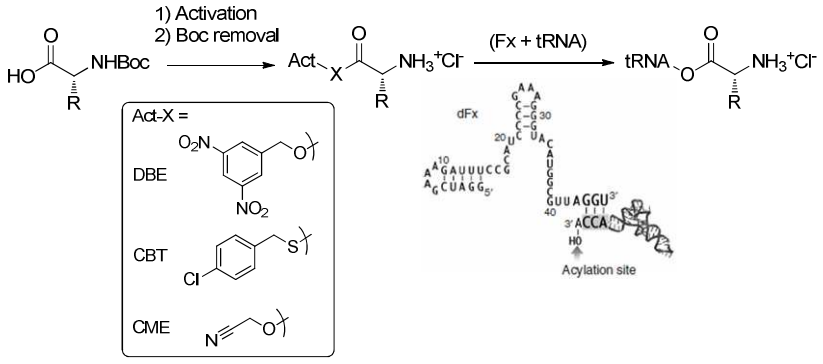


Figure 3. General scheme of Flexizyme mediated acylation of tRNA. Activation step is done by S_N2 reaction between carboxylic acid and 3,5-dinitrobenzyl-chloride (DBE-Cl) or chloroacetonitrile (CME-Cl) or by EDC-coupling between carboxylic acid and chlorobenzyl-thiol (CBT).

The advantage of this methodology is its easy implementation. Chemical synthesis of activated amino acid is straight forward. The disadvantage is that amino acid acylation condition, i.e. time, need to be individually optimized also the one need to prepare each AA-tRNA

separately. We prepared several activated D-amino acid according to the scheme above and have successfully load them on to an amber suppressor tRNA.

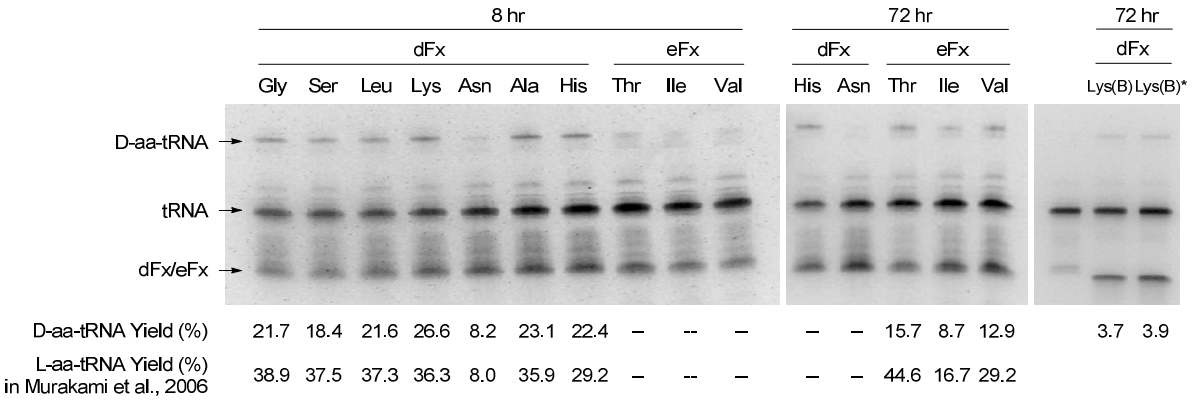


Figure 4. Analysis of flexizyme charged tRNA. After acylation reaction, the aa-tRNAs are labeled with 7.5 mg/mL of NHS-Biotin in 400 mM Hepes-KOH buffered at pH 8 and then analyzed on 15% TBE-Urea gel with 0.2 mg/mL streptavidin in loading buffer. Lys(B) denotes D-Lys(Biotin). *gel analysis without NHS-Biotin labeling.

Several standard amino acids work well with flexizymes, such as Gly, D-Ser, D-Leu, D-Lys, D-Ala and D-His, but there is also a few acylate poorly. Solubility of activated amino acid in the aqueous reaction environment is considered the cause of poor reactivity for hydrophobic residues (Murakami et al., 2006). This also explains why we observe very low acylation rate of the D-Lys(Biotin)-DBE substrate. Murakami et al. claims that the low observed yield by this assay is partly due to the deacylation during the NHS-Biotin labeling reaction. Earlier literature also stated that aa-tRNA deacylate very fast at pH > 6 conditions, with half-life ranged from few minutes to an hour depend on the side chain (Hentzen et al., 1972). To confirm if the low level of D-Lys(Biotin)-tRNA observed is due to poor acylation, we duplicate the lane without treatment of NHS-Biotin and analyze yield of acylation side-by-side with the NHS-Biotin treated sample. As it shows on the last two lanes of Figure 4, the observed yields are about the same with or without NHS-Biotin treatment, we conclude that the reaction itself is not efficient in this case. According to all published articles from Suga's group and personal communication with Dr.

Suga, we learned that besides extending reaction time, which we have done from 8 hours to 72 hours, there is no simple mean to further enhance efficiency.

In the next chapter we will describe ribosome mutants screening system which utilizes D-Lys(Bio) and several other non-proteingenic dialkylamino acid which also acylate very poorly by flexizyme system. Therefore, we also establish the chemical acylation in our lab as a complement for the flexizyme system with slight modification from (Ellman et al., 1991). As shown in Figure 5, the N-protected amino acid is first activated as cyanomethylester and then coupled to dinucleotide 'pdCpA' at either 2'-OH or 3'-OH. It is not necessary to isolate the 3'-O acylated product, since the intramolecule transacylation is rapid and EF-Tu would selectively pick up the 3'-O-aminoacyl tRNA (Sprinzl, 2006). The acylated dinucleotides are then ligated to *in vitro* transcribed tRNA without 3'–CA. We use NPPOC as N-terminus protecting group (PG) since its convenience to remove (illuminated by 100W hand-held UV-lamp) than traditional Nvoc (illuminated by dangerous 1000W Xenon lamp with filter) and 4-pentenoyl (I_2 / pyridine followed by ethanol precipitation).

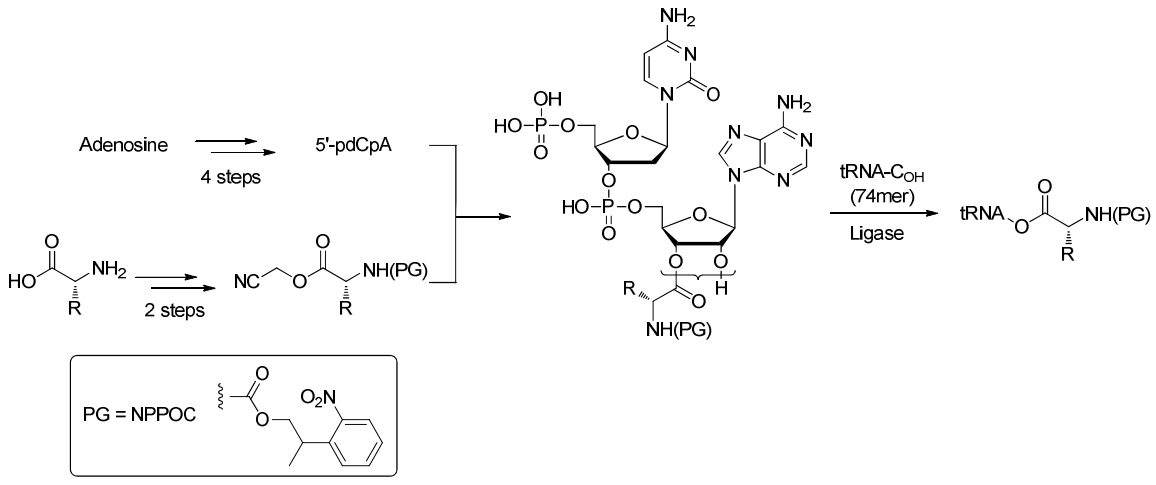


Figure 5. Scheme of chemical acylation. N-protected amino acid is first activated as cyanomethylester and then coupled to dinucleotide 'pdCpA' at either 2'-OH or 3'-OH. The acylated dinucleotides are then ligated to *in vitro* transcribed tRNA without 3'–CA.

We have found that the NPPOC protected aminoacyl dinucleotides are very stable at pH 7.4 buffered solutions (no hydrolyzed pdCpA detected), and the T4 RNA ligase mediated ligation is generally efficient disregarding what amino acid attached. To our surprise, when same *in vitro* transcribed tRNAs loaded with same amino acid by each method are subjected to a protein synthesis assay, we found the chemical acylated tRNA leads to better yield. Potential explanation is that the NPPOC protecting group removed right before translation reaction also protects the aa-tRNA from hydrolysis during sample storage and handling. In summary, we have built and tested several activated d-amino acid substrates for flexizyme acylation and several D-aa or α,α-dialkyl-aa-pdCpA for ligation method. They are listed in the table below:

via Flexizyme		**via Ligation**	
std. AA	**non std. AAs**	**std. AA**	**non std. AAs**
Gly-DBE	D-Lys(Biotin)-DBE	Gly-pdCpA	N-Biotin-L-Nvl-pdCpA
D-Ser-DBE	Aib-DBE	L-Ile-pdCpA	N-Ac-L-Lys(Biotin)-pdCpA
D-Leu-DBE	D-Iva-DBE	L-Ala-pdCpA	L-Dap(DEAC)-pdCpA
D-Lys-DBE	D-Phe(N_3)-CME	D-Ala-pdCpA	D-Dap(DEAC)-pdCpA
D-Ala-DBE		L-Val-pdCpA	Aib-pdCpA
D-His-DBE		D-Val$^{1\text{-}^{13}C}$-pdCpA	Deg-pdCpA
D-Asn-DBE			D-Iva-pdCpA
D-Pro-DBE			
L-Leu-DBE			
D-Thr-CBT			
D-Ile-CBT			
D-Val-CBT			
D-Phe-CME			
L-Phe-CME			

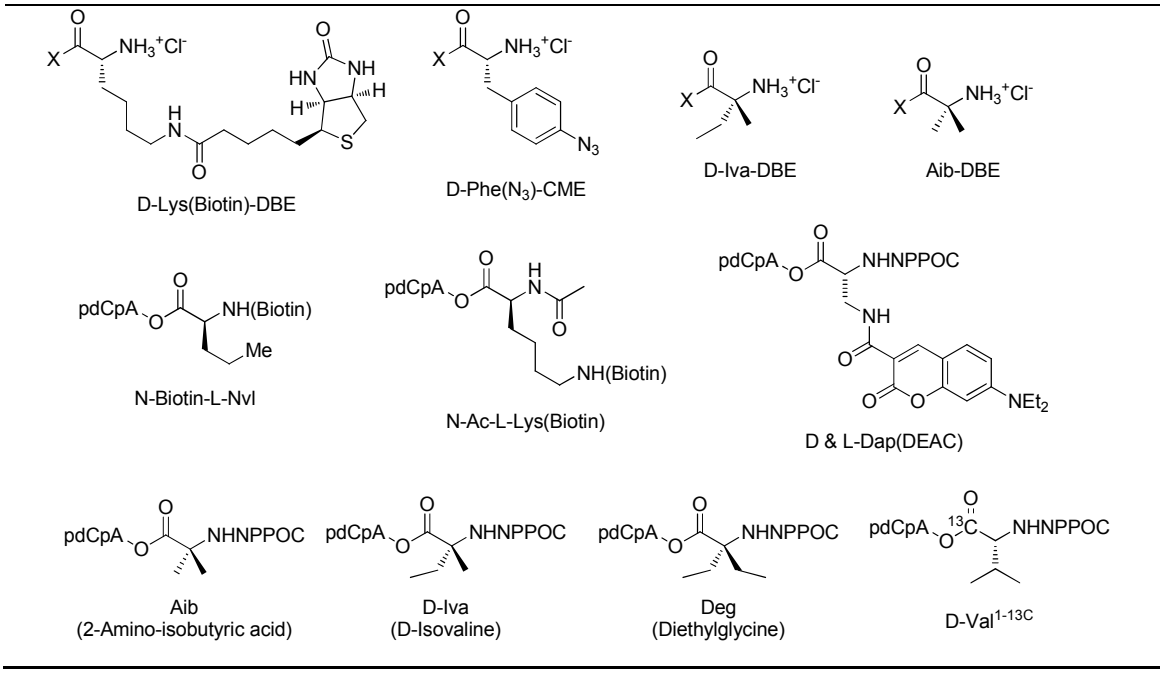


Table 2. Amino acid building blocks synthesized for mis-acylation in this project. Structures of non-standard amino acid are shown.

Two assays for D-amino acid incorporation and the interaction of aa-tRNA and EF-Tu

While we building a ribosome selecting system looking for better D-amino acid incorporation, we attempted to measure whether the D-aa incorporation through fluorescence gel imaging by the use of L-Lys(BODIPY)-tRNA^{Lys} (commercially available) or D/L-Dap(DEAC)-tRNA but with no success. Difficulty of incorporating bulky dye moieties is causing low protein production already. In order to have identifiable and amplifiable signal of the expressed peptide containing *in vitro* misacylated amino acid, we use western blot to detect the expression of a downstream FLAG-epitope after read-through D-aa-tRNA corresponding codon (Figure 6a). As for first screening for experiment conditions, this method is sensitive and handy, but with potential caveat that the residue incorporated is not the amino acid supplied on aa-tRNA. In order to reduce chance of mis-recognition of endogenous aaRS to supplied aa-tRNA, we reduce the template composition and hence the use of aaRSs and amino acids in PURE translation to

only six amino acids: Asp (D), Tyr (Y), Lys (K), Trp (W), Met(M) and Leu(L)[1]. The templates we designed contain N-terminus FLAG-tag coding sequence followed by one or two consecutive unassigned codons (amber stop codon without RF1 or amino acid codon without aaRS) and then a coded stretch of 78-amino acid containing mixed D, Y, K, L, M and W in order to stick read-through peptide on to blotting membrane.

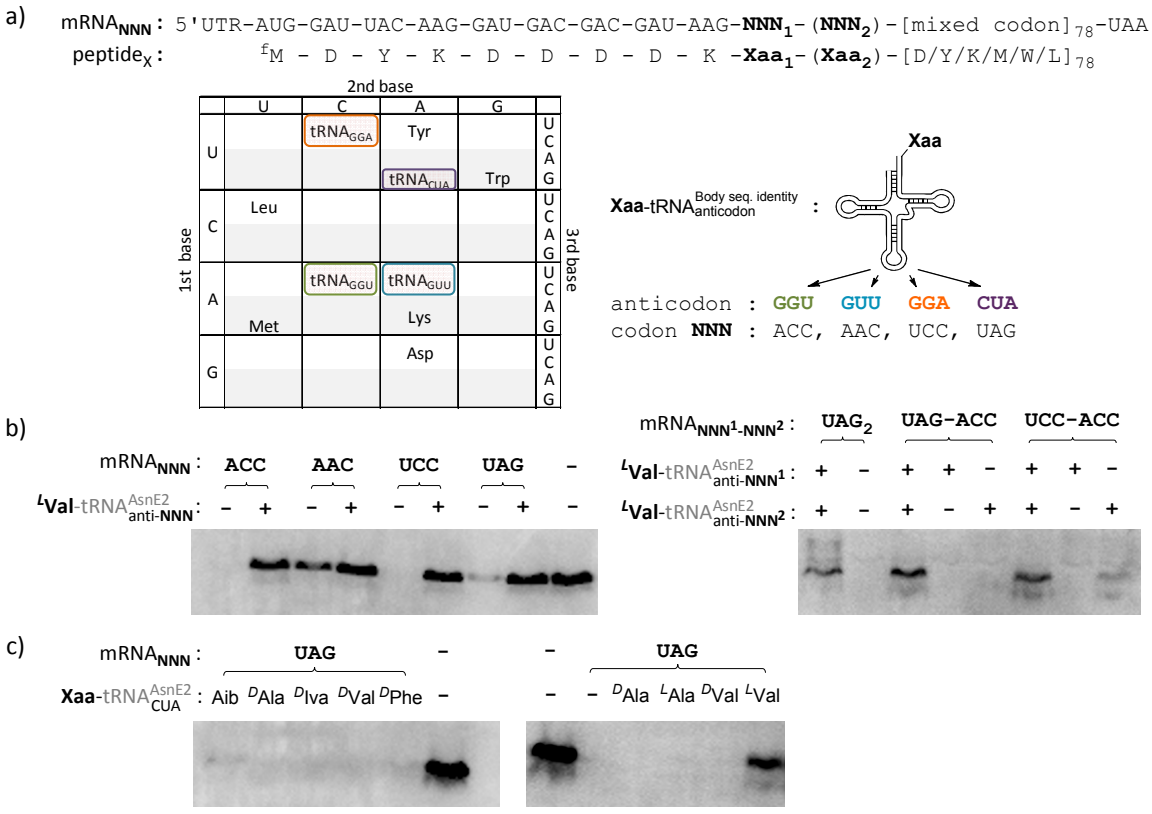


Figure 6. (a) General read-through assay used in this study, peptide with N-term FLAG-tag is expressed followed by one or two unassigned codon NNN and then a stretch of 78mer DYKWLM amino acids. The four anticodons-codon pairs are shown on the right and each $tRNA_{anticodon}$'s potential decoding range (including wobble) is shown on the left in codon table. (b) Test of aa-$tRNA_{anticodon}$ dependent read-through for four codons ACC, AAC, UCC, UAG and three consecutive dual codons UAG-UAG, UAG-ACC and UCC-ACC. The presence of band of FLAG-containing peptide indicate successful read-through of codons by chemically acylated tRNAs. (c) Test incorporation of several D, L-amino acids. mRNA with label "–" is a template without middle NNN codon and hence full-length peptide synthesis doesn't require chemically acylated

[1] Asp(D), Tyr(Y) and Lys(K) are chosen since they are required for FLAG-tag; Met(M) is needed for translation initiation; Trp(W) has intrinsic fluorescence which, although weak, will be used as trace in HPLC purification example later; Leu(L) is added to adjust overall peptide hydrophobicity, as well as to reduce DNA coding repetition.

tRNA. $tRNA^{AsnE2}$ is evolved from E. coli $tRNA^{Asn}$ to be orthogonal from all 20 aaRSs (Goto et al., 2011). All experiments shown in this figure are conducted with12 μM of misacylated aa-tRNAs (prepared by chemical acylation) and 2 μM of EF-Tu.

To find out optimum condition to measure unnatural amino acid (Uaa) incorporation rate, we began with the choice of codon for single and consecutive dual incorporation. It has been pointed out that the "fourth nucleotide", i.e. the first nucleotide in the next codon, has significant effect on decoding efficiency for UAG suppressor tRNAs in general, with better efficiency for codon followed by purines and worse for one followed by pyrimidines (Bossi, 1983; Smolskaya et al., 2013). However, although findings from Ayer et al. (Ayer and Yarus, 1986) against hypothetical base-pairing of suppressor tRNA U33 to the fourth codon, it suggests that these bias are less significant when relatively more efficient suppressors are used. We pick four codons ACC, AAC, UCC and UAG to test, and transplant their corresponding anti-codon into an orthogonal tRNA evolved by Suga's group (Goto et al., 2011). The translation reaction is carried out with only six amino acids, aaRSs and deacylated *E. coli* total tRNA. On the left panel of Figure 6b, the read-through of single ACC, UCC and UAG are mainly depend on the supplied aa-tRNA, and their decoding efficiencies are at the same level. Since there is significant background read-through of AAC codon, we leave it out of our candidate list. Notice that the UAG codon also has low but not negligible background read-through. Therefore, in later section I will describe a 2nd assay we developed to verify the identity of the incorporated residues to prevent misinterpretation. On the right panel of Figure 6b, when testing for two consecutive read-through events, the overall decoding yield of UAG-AAC surpasses the yield of UAG-UAG, agreed with the codon context effect mentioned above. The decoding of UCC-ACC is in between UAG-UAG and UAG-ACC, but to our surprise, it seems that aa-$tRNA_{GGU}$ targeting the ACC codon can also read UCC and yield full-length peptides when aa-$tRNA_{GGA}$ is not present. For the sake of simplicity, we use only UAG-UAG and UAG-ACC in downstream experiment. When

we applied this assay to a few D- or L-aas, we could only see incorporation of L-Val. Although L-Val-tRNA has one of the longest half-life toward hydrolysis in neutral pH solution, i.e. 60 mins versus Ala-tRNA's 6 mins (Hentzen et al., 1972), we found neither increasing initial concentration of L-Ala-tRNA nor adding multiple "booster doses" of L-Ala-tRNA during the course of translation helps to improve yield significantly.

This finding leads us to ask that whether the difficulty of incorporating L-Ala is the result of its aa-tRNA's poor interaction with EF-Tu. We also concerned if there is a competitive binding of aa-tRNAs resulting from the limiting amount of EF-Tu in the PURE translation system we used. *In vitro* experiment (Asahara and Uhlenbeck, 2002; Louie et al., 1984) shows that 21 different *E. coli* unmodified tRNA body sequences, when all acylated with same amino acid such as L-Val, display a wide range of binding affinities to EF-Tu from K_D = 0.5 nM for ^{L}Val-tRNA^{Glu} to 310 nM for ^{L}Val-tRNA^{Tyr}. Also, this trend is modulated by the corresponding acylated amino acid, i.e. the weaker binder tRNA^{Tyr} are linked to stronger binding promoter L-Tyr and *vice versa* for tRNA^{Glu}, in order to have all endogenous aa-tRNA shared similar affinity to EF-Tu. One the other hand, when the same tRNA is loaded with different L-amino acids the range of affinity toward EF-Tu spans at least 80-fold (Dale et al., 2004). The list of experimental dissociation constants in these literatures are reorganized in the table below:

Xaa-tRNA$^{Body\ seq.}$	K_D [nM]	Val-tRNA$^{Body\ seq.}$	K_D [nM]	ΔG° [kcal/mol]	Xaa-tRNA$_{NNN(Xaa)}$	K_D [nM]	ΔG° [kcal/mol]
Glu-tRNA^{Glu2}	43.6	Val-tRNA^{Glu2*}	0.5	-11.7			
Asp-tRNA^{Asp1}	28.3	Val-tRNA^{Asp1*}	1.9	-11.0	Asp-tRNA^{YFA2}	>150.0	< -8.6
Gly-tRNA^{Gly3}	11.4	Val-tRNA^{Gly3}	2.8	-10.7	Gly-tRNA^{YFA2}	62.0	-9.1
Thr-tRNA^{Thr}	16.1	Val-tRNA^{Thr3}	4.0	-10.5	Thr-tRNA^{YFA2}	15.0	-9.9
Ala-tRNA^{Ala1B}	28.6	Val-tRNA^{Ala2}	4.3	-10.5	Ala-tRNA^{YFA2}	100.0	-8.8
Cys-tRNA^{Cys}	13.6	Val-tRNA^{Cys}	21.0	-9.6			
Leu-tRNA^{Leu4}	24.7	Val-tRNA^{Leu1}	23.0	-9.5			
Met-tRNA^{Met}	10.6	Val-tRNA^{Met}	33.0	-9.4	Met-tRNA^{YFA2}	17.0	-9.8
Pro-tRNA^{Pro}	12.6	Val-tRNA^{Pro3}	34.0	-9.3	Pro-tRNA^{YFA2}	15.0	-9.8
Phe-tRNA^{Phe}	13.8	Val-tRNA^{Phe}	48.0	-9.2	Phe-tRNA^{YFA2}	11.0	-10.1
Lys-tRNA^{Lys1}	43.5	Val-tRNA^{Lys}	53.0	-9.1	Lys-tRNA^{YFA2}	35.0	-9.4
Arg-tRNA^{Arg2}	26.5	Val-tRNA^{Arg2}	54.0	-9.1	Arg-tRNA^{YFA2}	17.0	-9.8

Ser-tRNA^{Ser}	15.9	Val-tRNA^{Ser1*}	61.3	-9.1			
Asn-tRNA^{Asn}	10.6	Val-tRNA^{Asn*}	88.3	-8.9			
Val-tRNA^{Val1}	**92.0**	Val-tRNA^{Val1}	92.0	-8.8	Val-tRNA^{YFA2}	17.0	-9.8
Ile-tRNA^{Ile}	27.0	Val-tRNA^{Ile1}	110.0	-8.7	Ile-tRNA^{YFA2}	8.1	-10.3
Trp-tRNA^{Trp}	9.9	Val-tRNA^{Trp*}	183.6	-8.5	Trp-tRNA^{YFA2}	3.6	-10.7
Gln-tRNA^{Gln}	5.7	Val-tRNA^{Gln2}	250.0	-8.3	Gln-tRNA^{YFA2}	1.9	-11.1
Tyr-tRNA^{Tyr2}	15.7	Val-tRNA^{Tyr2}	310.0	-8.1			
					^{L}Tyr-tRNA^{Tyr}	50.0	-9.2
					^{D}Tyr-tRNA^{Tyr}	1200.0	-7.5

Table 3. Dissociation constant of aa-tRNA and EF-Tu•GTP complex (column 4, 5, 7, 8) were re-presented from (Asahara and Uhlenbeck, 2002; Dale et al., 2004); L- or D-Tyr-tRNA^{Tyr} data are from (Yamane et al., 1981). For Xaa-tRNA$^{Body\ seq.}$ (column 2), K_D of Phe and Leu are obtained first from the relative K_D ratio of Val/Phe and Val/Leu4 from (Louie et al., 1984), and then we refer to K_D of Phe and Leu for all rest L-aas. When same amino acids are acylated to different tRNA body sequences, the range of K_D spans ca. 600-fold, and when various amino acids are acylated to same tRNA, the range of K_D spans at least 80-fold based on the 13 tested amino acids. The L-Tyr acylated tRNA^{Tyr} binds to EF-Tu•GTP 25-fold stronger than D-aminoacylated version. Notice that the dissociation constants are measured at 0 °C, the same constant at 37 °C is 6-10 folds higher, calculated using $\Delta G° = -RT \ln (1/K_D)$. *Data converted from the measured K_D of Phe-tRNA$^{Body\ seq.}$.

In our experiment, the tRNA^{AsnE2} seems to be one of the very poor binders as its ancestor tRNA^{Asn}, since most of the residues in contacted with EF-Tu are shared between the two. If so, plus if the total amount of endogenous aa-tRNAs exceeds available EF-Tu during *in vitro* translation, then even in the case of L-amino acid, linearly increase of the concentration of misacylated aa^{weak}-tRNA^{weak} will have only little to negligible effect on equilibrium concentration of ternary complex [EF-Tu•GTP•aa-tRNA]. This is similar to what we saw. To investigate whether our PURE translation set up is sub-optimized, we summarized the relative amount of EF-Tu and level of aaRS-maintained aa-tRNAs in literatures which is demonstrating unnatural amino acid incorporation in cell free protein synthesis below (Table 4):

Year, Lab	**Translation system**	**[EF-Tu],** μM	**[aa-tRNA^{aaRS}],** μM	**[aa^{mis}-tRNA],** μM
–	*E. coli* cell (mid-log phase)	ca. 100*[a]	ca. 100*[b]	–
1992-2013 Schultz P. & Hecht S.	Cell extract	ca. 25*[c]	ca. 6.4	12
2001-2013 Ueda T.	Purified comp. (all 20 aaRS)	2*[d]	ca. 53	20
2007-2013 Suga H.	Purified comp. (with 4 aaRS)	10*[e]	ca. 13*[f]	25-600

2014 This study	Purified comp. (with 6 aaRS)	2, 25*[g]	ca. 18*[f]	12-24

Table 4. Comparing the relative amount of EF-Tu and level of aaRS-maintained aa-tRNAs in literatures and our assay. *[a] ref. from (Pedersen et al., 1978); *[b] ref. of total tRNA concentration is from (Dong et al., 1996), and ref. of total aa-tRNA concentration is from (Jakubowski and Goldman, 1984); *[c] ref. from (Ellman et al., 1991); *[d] ref. from (Shimizu and Ueda, 2010); *[e] ref. from (Goto et al., 2011); *[f] calculated based on total tRNA used in reaction times fraction of isoacceptors will be charged by the supplied aaRS; *[g] 2 μM EF-Tu is used in experiment shown in Figure 6, 25 μM is used in all subsequent experiments after optimization.

In our previous assay shown in Figure 6, we used 2 μM of EF-Tu and 53 μM of tRNA (ca. 18 μM been acylated by the 6 supplied aaRS) in the translation system based on Ueda's protocol, which now seems to be an adverse condition for incorporating ^{L}Ala-tRNA, an acyl-ester that promotes aa-tRNA•EF-Tu interaction poorly. In fact, L-Ala acyl ester might be worse in promoting EF-Tu association than some good ones in D-form such as D-Tyr, if assuming L-to-D change will invariably reduce affinity by 25-fold. This suggests that the previously perceived trouble on incorporating D-aas might to a great extend be the same one troubling ^{L}Ala here. Therefore, we examine if increasing either [EF-Tu], [^{L}Ala-tRNA^{AsnE2}] or both could alleviate the problem. We first investigate the effect of [EF-Tu] versus peptide production without involving Xaa-tRNA (Figure 7a) to find highest tolerable [EF-Tu]. We found that the best expression is achieved between 10-25 μM of EF-Tu, and the yield drops above that. However, at 35 μM of EF-Tu one could still get ca 84% of yield comparing to the maximum yield.

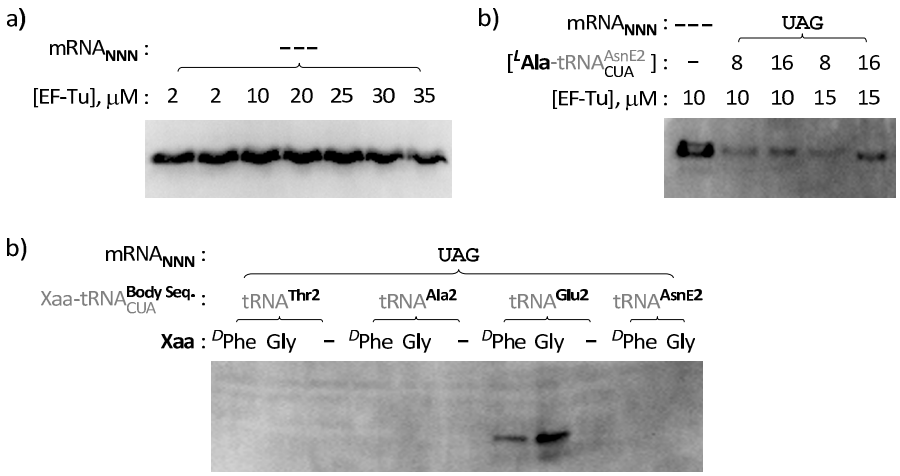


Figure 7. (a) Protein synthesis versus [EF-Tu] used, in searching for the highest tolerable [EF-Tu]. (b) Increasing either [^{L}Ala-tRNA^{AsnE2}] or [EF-Tu] or both improves incorporation rate. Except when [^{L}Ala-tRNA^{AsnE2}] is as low as 8 μM, increasing [EF-Tu] seems to have no effect. Around 13 μM of endogenous tRNA are used. (c) Various tRNA backbones with CUA anticodon are tested for the incorporation of Gly and ^{D}Phe, with 24 μM of chemical acylated aa-tRNA and 25 μM of EF-Tu. Good tRNA body such as Glu2 allows observable ^{D}AA incorporation. mRNA with label "---" is a template without middle NNN codon and hence full-length peptide synthesis doesn't require chemically acylated tRNA.

In Figure 7b, we run reactions with 12 μM of total aaRS-maintained [aa-tRNA], and test whether or not the excess EF-Tu (15μM EF-Tu minus 13 μM for endogenous aa-tRNA and 2 μM left) would promote incorporation by raising EF-Tu-bound ^{L}Ala-tRNA^{AsnE2} in equilibrium. The result is somewhat inclusive that while increasing [EF-Tu] does improve yield when 16 μM of aa-tRNA is used, it has no obvious effect when 8 μM of aa-tRNA is used. But nonetheless, if increasing each's concentration of the two associating partners could both promote aa-tRNA delivery, then it is reasonable to expect that tuning up the affinity between the two would do the same trick.

Accordingly, we build tRNAs with body sequence taken from the top list of strong binders[2], and test if they were able to deliver non-preferred amino acids better. Although

[2] The list on Table 3 includes Thr2, Ala2, Glu2, Asp1 and Gly3. We didn't take Asp1 since we have AspRS during translation, and we use GlyRS to make peptide for the HPLC assay which will be described later.

orthogonality of $tRNA^{AsnE2}$ toward all *E. coli* aaRSs is usually a desired property, this is not likely to trouble us since we only use 6 aaRS at this stage. In the long term, more Xaa would be brought in by non-enzymatic acylation method in order to use as much codons as possible, and aaRS would be slowly phased out eventually. As shown in Figure 7c, when these tRNAs are acylated with Gly, which is as poor as Ala in promoting tRNA-EF-Tu binding, ^{D}Phe or used as free 3'-OH, we found the Glu2 backbone shows superior capacity to incorporate both Gly and ^{D}Phe while none of the other tRNAs can do either one.

Mutagenesis study of EF-Tu for enhanced D-AA delivery capacity

The striking bias of EF-Tu on tRNA backbones in the incorporation of chemically charged aa-tRNA reminds us that it could also play role as significant as ribosome in discriminating D-amino acid. The previous biochemical data has shown its 25-fold preference for ^{L}Tyr versus ^{D}Tyr-$tRNA^{Tyr}$, and the binding of ^{D}Tyr-$tRNA^{Tyr}$ to EF-Tu doesn't protect it from hydrolysis, which indicates ^{D}Tyr-acyl ester must adopt a different conformation than L-form when bound. In crystal structure (PDB ID: 1OB2) of EF-Tu in ternary complex with Phe-$tRNA^{Phe}$, the region surrounding amino acyl Cα-hydrogen, where D-side chains would locate, is well packed, which suggests there is a steric determinant applied for all amino acid substrates. To avoid complication on considering how each D-side chains of Daa-tRNA would affect their association with EF-Tu, and also in attempt to restore EF-Tu's ability to protect Daa-acyl ester from hydrolysis, a preliminary mutagenesis study was performed on EF-Tu.

We pick eight residues on EF-Tu that are closest to the D-side Cα proton of aa-tRNA, in an 'alanine scanning' fashion, each mutant contains one bulky residue mutated to alanine among these eight sites, with the hope to relax the packing in this region. Doi Y. et al. has applied the

same strategy to successfully create mutants capable to tolerant amino acid with bulky side chains such as L-1-Pyrenylalanine or DL-2-anthraquinonylalanine (Doi et al., 2007). These residues are D216, V226, V227, T228, V258, N273, V276 and L277.

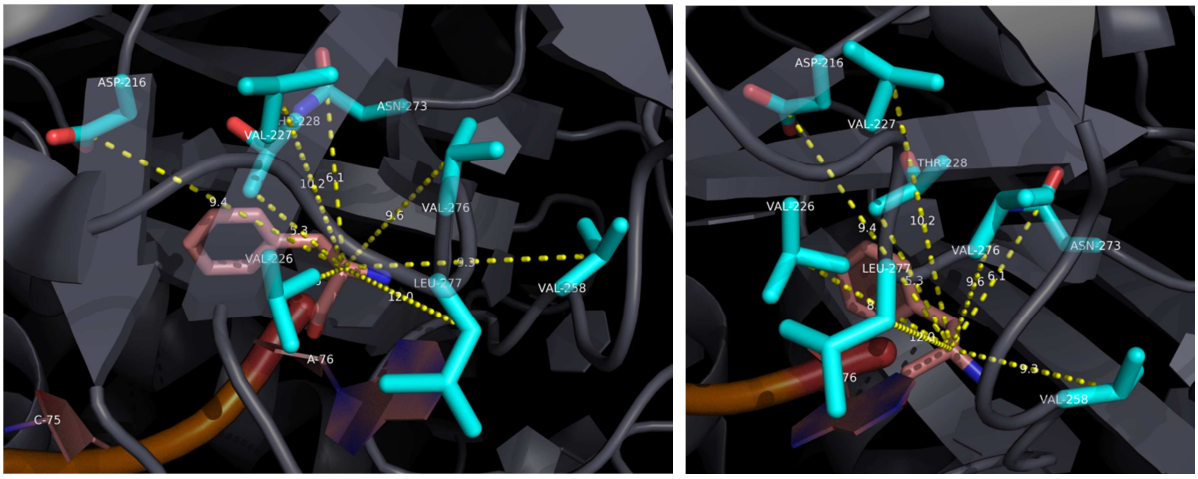


Figure 8. Illustration of eight EF-Tu residues around C_α-H of ^{L}Phe-tRNA been mutated to Ala in this study. Cartoon in grey is EF-Tu; side-chains of residues been mutated are plot as stick in cyan; Phe is plotted as stick in pink. Structure is generated based on PDB ID: 1OB2 (unpublished data).

These mutants are cloned and purified with C-terminus His_6-tag, as the wild-type. When their activities on expressing normal peptide are tested, to our surprise, all mutants promote translation similarly (Figure 9a). When assaying their abilities to incorporate artificially charged aa-tRNAs, mutant V276A shows significant improvement on yield while others perform the same as wild type. It is interesting that V276 has no direct contact with amino acid on aa-tRNA (Figure 9c), and it not only promotes D-aa incorporation but also L-aa such as Gly. Based on the crystal structure, V276 is not the residue (9.6 Å) having direct contact with aa-tRNA among the 8 residues mutated in this study, this effect is unexpected, as it suggests that there is still room to improve on efficiency of the aa-tRNA•EF-Tu•ribosome system although it has been naturally evolved for 3.5 billion years since the cyanobacteria. We are building more mutants with more Ala substitutions around V276 and also screen the region surrounding the L-side chain of

aminoacyl ester instead of D-side, such as H66, F261 and M260, since (Yamane et al., 1981) proposed D-aminoacyl ester bind to EF-Tu with different conformation than L-form.

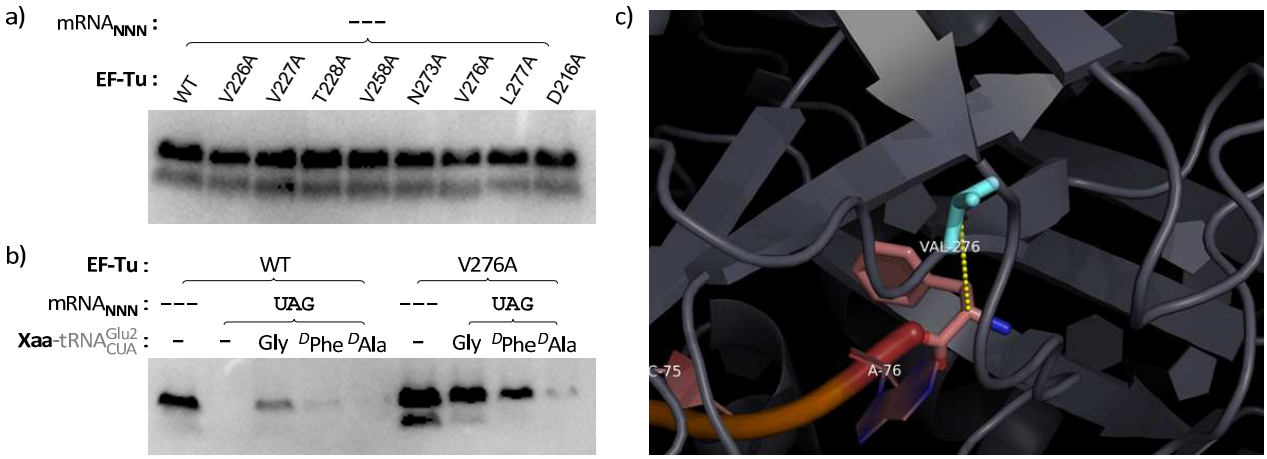


Figure 9. (a) Assay translation activities of 8 EF-Tu mutants on normal peptide. All mutants perform at similar level as wild-type. (b) Comparing wild-type EF-Tu and mutant V276A on their capacity to promote incorporation of artificially charged aa-tRNAs. mRNA with label "---" is a template without middle NNN codon and hence full-length peptide synthesis doesn't require chemically acylated tRNA. (c) Distance between EF-Tu residue V276 to Cα of ^{L}Phe-tRNA and their context. Structure is generated based on PDB ID: 1OB2 (unpublished data).

HPLC assay to validate D-amino acid incorporation

Western blot of D-aa-tRNA dependent read-through peptide expression is the first evidence we observed. In order to confirm it is D-amino acid, not trace amount of enantiomeric L-form impurity instead, we design a second assay based on HPLC to check their identity. In this assay, as shown in Figure 10a below, $mRNA_{UAG\text{-}ACC}$ containing two unassigned codons is used to express $peptide_{XY}$, which is isolated by FLAG affinity purification and further clean up by HPLC. They are then acid hydrolyzed to yield individual amino acids[3], derivatized by a chiral auxiliary group (+)-1-(9-Fluorenyl)ethylchloroformate (FLEC) and analyzed by HPLC. We

[3] To prevent complication of signal and to reduce noise, we didn't add any additives in 6N HCl during hydrolysis such as phenol or thiol. Therefore Met, Tyr and Trp are not recovered, as they will be oxidized during hydrolysis.

applied this method to ^{D}Ala and ^{D}Phe by putting them onto tRNA corresponding to UAG codon. After translation and FLAG affinity purification, on HPLC, we found that peptides with or without Xaa incorporation have different retention time, which is small in the case of D/L-Ala but larger for D-Phe. We collect these fractions and performed hydrolysis and derivitization. HPLC trace of labelled amino acid hydrolysates from peptide$_{L\text{-Ala}}$, peptide$_{D\text{-Ala}}$ are compared with labelled standard L-amino acids plus ^{D}Ala and ^{D}Phe. Contrasting the trace of peptide$_{L\text{-Ala}}$, peptide$_{D\text{-Ala}}$, ^{D}Ala signal is only presented in peptide$_{D\text{-Ala}}$ hydrolysate, which indicates that it comes from the peptide expressed but not from racemization of ^{L}Ala during acid hydrolysis.

Unfortunately, the HPLC analysis of peptide$_{D\text{-Phe}}$ is inconclusive due to overwhelming protein contamination, so we are currently looking for additional purification strategy to resolve this issue.

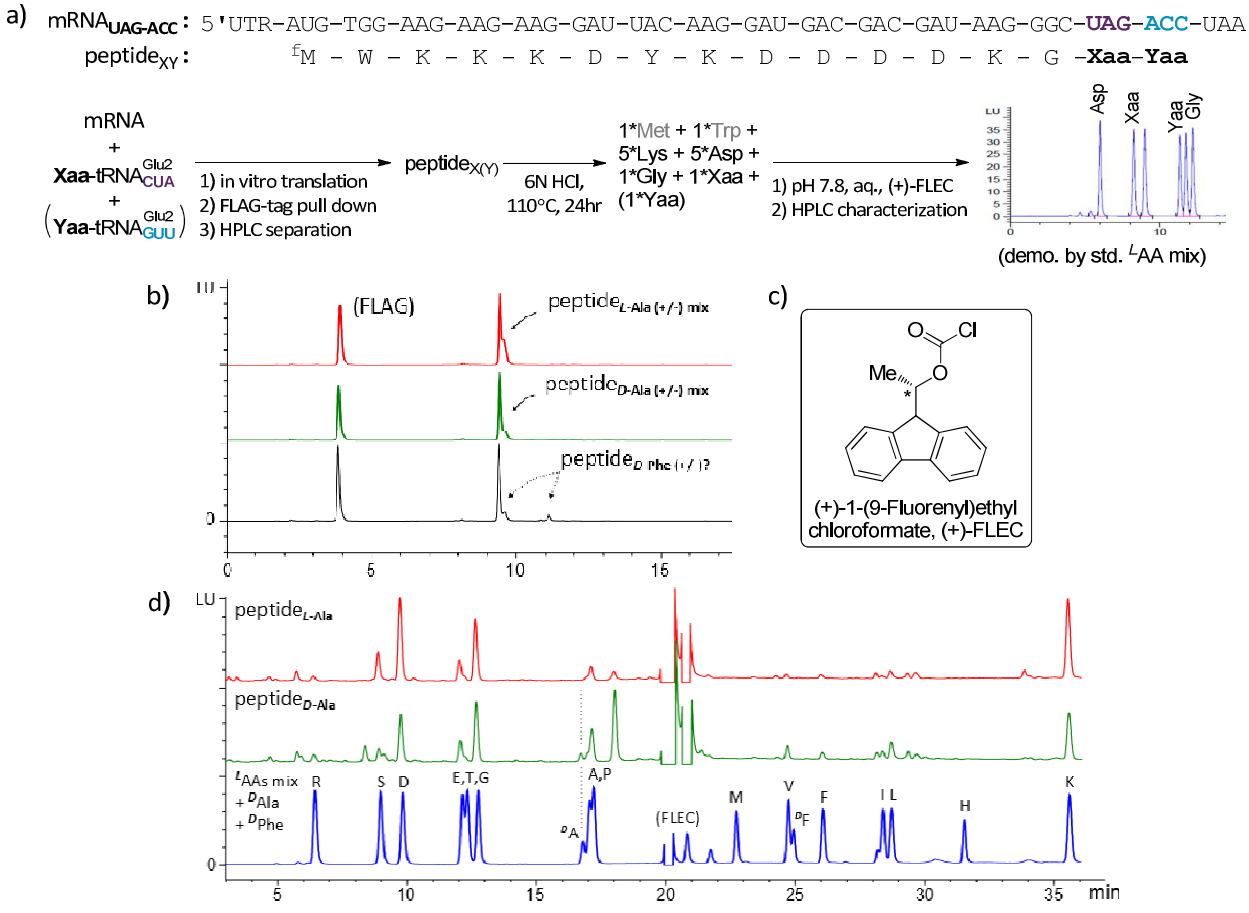


Figure 10. HPLC assay of incorporated amino acid. (a) Template contains codons assigned to artificially charged aa-tRNA are placed after FLAG-tag for affinity pull down. The eluted peptides are then purified by HPLC and hydrolysis in 6N HCl to individual amino acids. Amino acids are then derivatized by chiral auxiliary (+)-1-(9-Fluorenyl)ethylchloroformate and analyzed by HPLC with standard C18 column. Notice we utilize only UAG codon in this experiment here. (b) HPLC trace of three affinity purified peptides. The peak on the left corresponds to FLAG peptide used for elution. Peptides with incorporated D/L-Ala or D-Phe might share close retention time to those from peptides without any incorporation. (c) Structure of (+)-1-(9-Fluorenyl)ethylchloroformate (d) HPLC trace of (+)-FLEC labelled amino acid hydrolysate from peptide$_{L\text{-Ala}}$ and peptide$_{D\text{-Ala}}$. Trace of standard amino acid mixture plus ^{D}Ala and ^{D}Phe are shown on the bottom.

Discussion

In this chapter we discussed two factors discriminating D-amino acid from L-amino acid in bacterial translation system, amino acyl tRNA synthetases (aaRS) and elongation factor Tu (EF-Tu). We began with the survey on the substrate specificities of all 19 L-aaRSs, in order to assess the possibility to use endogenous aaRS to load D-aa onto tRNA. The result confirmed the

literature report that only TyrRS, TrpRS and ArgRS shows low but measurable activity toward corresponding D-aa. Therefore we moved to chemical acylation and Flexizyme acylation, and several substrates are prepared for both strategies. When we first used these artificially charged tRNA to incorporate Xaa into peptide, we have problems even on some proteingenic L-aa such as ^{L}Ala. Therefore we took several approaches to address this issue all surrounding the same principle: increasing the affinity between EF-Tu and aa-tRNA. We tested if different codon affects yield, and found that it does not with the presence of single codon, but the yield is slightly compromised when successive dual UAG codon appears. We tested if tRNA backbone sequences affects yields, and it did. And we also found an EF-Tu mutant V276 that could further enhance the overall translation yield, with mechanism to be investigated.

The aa-tRNAs used for translation in this chapter are all chemically acylated, since it is operationally convenient once we have aa-pdCpAs available. The *in vitro* transcribed tRNA used for ligation has a 5'-GTP structure instead of the native 5'-GMP. We have compared the effect of 5'-GTP versus 5'-GMP and found that it does not compromise incorporation yield if not facilitate the process (data not shown). One possible explanation is that 5'-GMP tRNA complicates ligation reaction, and so following-up experiment is proposed to reexamine this effect with flexizyme charged tRNA.

To our surprise is that at lower concentration of ^{L}Ala-tRNA^{AsnE2} (8 μM) and EF-Tu (10 μM), increasing [aa-tRNA] has more prominent effect then increasing EF-Tu. Possible explanation is that increasing amount of [EF-Tu] increases EF-Tu•GTP ternary complex formation of other weak-binding aa-tRNA, therefore the overall probability of delivery to ribosome is not increased much.

However, we found that changing tRNA backbone sequences has more prominent affects than tuning concentration of each component. Non-preferred amino acids, such as Gly, ^{L}Ala or ^{D}Phe, could acquire moderate incorporation when acylated to a preferred tRNA backbone. However, when considering to incorporate more than one artificially charged amino acids with very different affinity toward EF-Tu, i.e. Xaa and Yaa, adjustment on the ratio of [Xaa-tRNA]/[Yaa-tRNA] is necessary to balance each ternary complex's concentration. Otherwise the weaker aminoacyl-tRNA would incorporate poorly. High concentration of the aa-tRNA with lower affinity is required to form adequate amount of ternary complex, since raising [EF-Tu] too much would hurt translation yield, as shown previously. But this is very inefficient and wasteful since the vast majority of weak aa-tRNA will be unbound, and they hydrolyze quickly comparing to time scale of translation reaction. Given that all other D-amino acids have lower incorporation rates than ^{D}Phe (Fujino et al., 2013), it would be desired to have EF-Tu tolerating D-aa as well as L-aa. And then pairing more preferable tRNA backbone with weak D-aa could be employed along with Glu2 to modulate more than one D-amino acids, without using large excess of particular aa-tRNA wastefully.

Materials and Methods

Preparation of amino acyl-tRNAs

Procedures to synthesis of CME-, CBT- and DBE-activated amino acids are adopted from (Murakami et al., 2006). Procedure for the synthesis of pdCpA are from (Ellman et al., 1991), NPPOC protection from (Bhushan et al., 2003). Identity and purity of all substrates are confirmed by either HRES-MS, NMR or both.

In vitro transcriptions of tRNA, dFx or eFx are done as follow: prepare 300 μL mixture of 4mM NTPs, 0.5 U/μL Murine RNase Inhibitor (New England Biolab Inc.), 10mM DTT, 100 μg/mL T7 RNA polymerase, 125 μg/mL inorganic pyrophosphatase, 1x T7 transcription buffer (contains 40mM Tris-HCl pH 8.1, 20mM $MgCl_2$, 0.1 % Triton X-100, 30mM Spermidine and 500 μg/mL BSA) and optional 10 % DMSO (for $tRNA^{Glu2}$ synthesis) and incubate at 42 °C for 2.5hr. After that, 200 μL mixture of same components except with 10mM NTPs, 30mM $MgCl_2$ and 20 μg/mL T7 RNA polymerase is added into previous reaction, and further incubate at 38 °C overnight. The product is washed with acid phenol twice and with 24:1 chloroform-isoamyl alcohol twice, and then precipitated with 500 μL of isopropanol. Crude RNAs are purified by 15% denaturing PAGE gel.

Ligation of aa-pdCpA to $tRNA_{C-3'}$ and Flexizyme charging: For a 200 μL reaction, mix the components on ice by the following order: 8 μL of 5 mM aa-pdCpA, 20 μL DMSO, 102 μL of water, 20 μL of 10x T4 ligation buffer (contains 500 mM Hepes-KOH pH 7.5, 150 mM $MgCl_2$ and 7.5 mM ATP), 20 μL of 400 μM $tRNA_{C-3'}$ and 30 μL of 10 kU/mL T4 RNA Ligase (New England Biolab Inc.). After incubate on 37 °C for 45 min, 30μL of 2M NaOAc pH 4.5 is added to quench reaction and N-protected aa-tRNA is recovered by EtOH precipitation and quantified by Qubit 2.0 Fluorometer. Right before use, chemical acylated aa-tRNAs are illuminated with 100W, 350 nm UV (B100-AP from UVP LLC.) for 10 min on ice to remove NPPOC protecting group. Flexizyme charging is done exactly as described in (Murakami et al., 2006)

<u>Protein purification and construction of PURE translation reaction</u>

PURE translation plasmids are kindly provided by T. Ueda. Overexpression and purification of aaRS and translation factors follows procedure from (Shimizu and Ueda, 2010) exactly. Ribosome is purified from strain RB1 (Wang et al., 2012) following same procedure as (Shimizu and Ueda, 2010). PURE translation master mix cocktail with limited aaRS and amino acids are prepared as follow. Solution A (5x) contains: 10 mM ATP, 10 mM GTP, 5mM CTP, 5 mM UTP, 100 mM phosphocreatine, 250 mM Hepes-KOH pH 7.60, 500 mM KOAc, 65 mM $Mg(OAc)_2$, 10 mM Spermidine, 5 mM DTT, 50 μg/mL formyl donor (Shimizu and Ueda, 2010). Solution B (10x) without EF-Tu contains: 12.12 μM IF1, 4.11 μM IF2, 4.86 μM IF3, 16.43μM EF-Ts, 6.44 μM EF-G, 1.21 μM AspRS, 0.86 μM GlyRS, 1.09μM LysRS, 0.28 μM MetRS, 0.29 μM TrpRS, 0.13 μM TyrRS, 5.85 μM fmt, 0.89 μM rabbit muscle creatine kinase, 1.15 μM yeast myokinase, 0.64 μM nucleotide diphosphate kinase, 1.01 μM T7 RNA polymerase, 0.46 μM inorganic pyrophosphatase. Translation reaction is mixed as follow: prepare solutions on ice containing 1x Solution A, 1.36 mg/mL deacylated total tRNA (MRE600 total tRNA from Roche, incubate in pH 8.0, 250 mM $NaBO_4$ buffer for 1 hr at 37 °C and then cleaned up by Zeba-desalting column), 0.1 mM of each amino acid (Met, Asp, Gly, Tyr, Trp, Leu and Lys), 1x Solution B, 1.2 μM ribosome, 25 μM EF-Tu, 4 ng/μL DNA template. Incubate mixture at 37 °C for 1 hr right after addition of 24 μM photo-deprotected aa-tRNA. The actual EF-Tu and aa-tRNA concentrations used are described in each figure, otherwise 25 μM of EF-Tu and 24 μM of aa-tRNA are used.

Western blot analysis of expressed peptides

DNA Templates for the read-through assay are custom synthesized (Integrated DNA Technology) with sequence below (bold: translation initiation site; lower case: FLAG tag; underscore: unassigned codon, ex TAG, AAC, ACC or UCC):

5’GGCGTAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAAC**ATG**gattacaa
ggatgacgacgataagNNNCTGTGGATGAAGAAAATGAAAAAGGACTGGAAGTATCTCGATT
GGGACATGGACATGATGGACTATTGGTGGATGGATGACCTGTGGCTGGATTACAAA
TGGGATGATCTTATGCTGATGGATAAGTACCTGGATGATATGGATGATGATTACTTG
ATGGATATGATGGACGATTGGGATCTCATGTTATGGTACCTCTACATGTATCTCCTG
GATGACTGGGATATGTATAAGTAA3’

Translation reactions are quenched with equal volume of 2x tricine SDS sample buffer (Life technology) and analyzed on 16 % Tricine protein gel. Proteins are transferred to PVDF membranes by iBlot® (Life technology, setting program = P3, duration = 3 min 40 sec), and then blotted by anti-FLAG M2 antibody (Sigma F1804) and detected by SuperSignal West Femto kit (Pierce) and imaged by Bio-rad ChemiDoc™ MP system.

Mutagenesis of EF-Tu

Cloning was based on the EF-Tu overexpressing plasmid constructed by T. Ueda’s group. To describe their construct in brief, *tufB* gene was added EcoRI and BglII site (stop codon removed, therefore His_6-tag coded on plasmid will be expressed) at N- and C-term and cloned into corresponding site of pQE60. Top and bottom strand primers which carry mutations and have 12bp 5’-overlap on designated mutated location are used to amplify the whole wild-type EF-Tu plasmid. The linearized products are then reassembly to plasmid by GeneArt® Seamless cloning kit (Life technology). Primers used in this study are list below:

Mutation	Top strand primer	Bottom strand primer
D216A	AAGCCGTTCCTGCTGCCGATCGAAGCGGTATTCTCCATC	CAGGAACGGCTTGTCAATCGCACGCTCTGGTTCC
V226A	GGTCGTGGTACCGCGGTTACCGGTCGTGTAGAA	GGTACCACGACCGGAGATGGAGAATACGTCTT
V227A	GGTCGTGGTACCGTTGCGACCGGTCGTGTAGAA	GGTACCACGACCGGAGATGGAGAATACGTCTT
T228A	GGTCGTGGTACCGTTGTTGCGGGTCGTGTAGAA	GGTACCACGACCGGAGATGGAGAATACGTCTT
V258A	ACCTGTACTGGCGCGGAAATGTTCCGC	GCCAGTACAGGTAGACTTCTGAGTCTCTTTGATACCA

N273A	CGTGCTGGTGAGGCGGTAGGTGTTCTGCTGCGTGGTATC	CTCACCAGCACGGCCTTCGTCCAGCAGTTTGC
V276A	CGTGCTGGTGAGAACGTAGGTGCGCTGCTGCGTGGTATC	CTCACCAGCACGGCCTTCGTCCAGCAGTTTGC
L277A	CGTGCTGGTGAGAACGTAGGTGTTGCGCTGCGTGGTATC	CTCACCAGCACGGCCTTCGTCCAGCAGTTTGC

Table 5. List of primers used in the mutagenesis EF-Tu.

Amino acid analysis of expressed peptides

Translation reactions are added 1/5 volume of dissociation mixture (contains 41.6 mM EDTA, 0.42 % Triton X-100 and 4.16M NaCl) and incubate on ice for 10 min. Peptide products from every 100 µL translation reaction mixture are bound to three successive batches of 10 µL anti-FLAG M2 magnetic beads (Sigma M8823) and washed with HST buffer (contains 5 mM Hepes-KOH pH 7.37, 150 mM NaCl, 0.05 % Tween-20) three times. After washed once more with HS buffer (HST buffer without Tween-20), the peptide products are eluted with 100µM FLAG peptide dissolved in HS buffer and then purified by HPLC (Equipment: Agilent 1260 series pump with fluorescence detector, column: Eclipse Plus C18 1.8 µm, 2.1 × 150 mm; buffer A: 0.1% TFA in water; buffer B: 0.1% TFA in acetonitrile; flow rate: 0.4 mL/min; gradient: 5% to 65% buffer B from 0 to 20min; FLD: excitation 277 nm, emission 355 nm). Fractions corresponding to peptide products are pooled and concentrated to dryness by speedvac and hydrolysis by 6N HCl vapor at 110°C under argon. Hydrolysates are resuspended in 25 µL buffers (containing 25% v/v acetonitrile and 0.2 M $NaBO_4$ buffer pH 7.8), and derivatized by mixing 25 µL of 0.9 mM (+)-1-(9-Fluorenyl)ethylchloroformate (FLEC). After incubating for 30 min at room temperature, the derivatizing reaction is quenched by the adding 5 µL of 1M KBO_4 pH 10.4 and the excess (+)-FLEC related impurities are removed by washing the mixture with 15 µL hexane 4 times. The aqueous phase is acidify by 5 µL 2M H_3PO_4 and analyzed by HPLC (column: Eclipse Plus C18 1.8 µm, 3.0 × 100 mm; buffer A: 8% acetonitrile, 12% tetrahydrofuran, 80% 30 mM NaOAc pH 4.38; buffer B: 40% tetrahydrofuran, 60% 30 mM

NaOAc pH 4.38; flow rate: 0.5 mL/min; gradient: 10% buffer B from 0 to 5 min, 10% to 30% buffer B from 5 to 13 min, from 30% to 100% solution B from 13 to 30min and at 100% solution B till 40min; FLD: excitation 263 nm, emission 313 nm)

References

Asahara, H., and Uhlenbeck, O.C. (2002). The tRNA Specificity of Thermus thermophilus EF-Tu. Proc. Natl. Acad. Sci. *99*, 3499–3504.

Ayer, D., and Yarus, M. (1986). The Context Effect Does not Require a Fourth Base Pair. Science (80-.). *231*, 393–395.

Bhushan, K.R., DeLisi, C., and Laursen, R.A. (2003). Synthesis of photolabile 2-(2-nitrophenyl)propyloxycarbonyl protected amino acids. Tetrahedron Lett. *44*, 8585–8588.

Björk, G.R. (1996). Stable RNA modification. In Escherichia Coli and Salmonella: Cellular and Molecular Biology, N. F.C., R. Curtiss III R., Ingraham J.L., Lin E.C.C., Low Jr K.B., Magasanik B., and S.M. and U.H.E. W.S., Riley M., eds. (Washington: ASM Press), pp. 861–886.

Bossi, L. (1983). Context effects: Translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message. J. Mol. Biol. *164*, 73–87.

Dale, T., Sanderson, L.E., and Uhlenbeck, O.C. (2004). The Affinity of Elongation Factor Tu for an Aminoacyl-tRNA Is Modulated by the Esterified Amino Acid†. Biochemistry *43*, 6159–6166.

Doi, Y., Ohtsuki, T., Shimizu, Y., Ueda, T., and Sisido, M. (2007). Elongation Factor Tu Mutants Expand Amino Acid Tolerance of Protein Biosynthesis System. J. Am. Chem. Soc. *129*, 14458–14462.

Dong, H., Nilsson, L., and Kurland, C.G. (1996). Co-variation of tRNA Abundance and Codon Usage inEscherichia coliat Different Growth Rates. J. Mol. Biol. *260*, 649–663.

Eigner, E.A., and Loftfield, R.B. (1974). [48] Kinetic techniques for the investigation of amino acid: tRNA ligases (aminoacyl-tRNA synthetases, amino acid activating enzymes). Methods Enzymol. *29*, 601–619.

Ellman, J., Mendel, D., Anthony-Cahill, S., Noren, C.J., and Schultz, P.G. (1991). [15] Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. Methods Enzymol. *202*, 301–336.

Freist, W., Wiedner, H., and Cramer, F. (1980). Chemically modified ATP derivatives for the study of aminoacyl-tRNA synthetases from baker's yeast: ATP analogs with fixed conformations or modified triphosphate chains in the aminoacylation reaction. Bioorg. Chem. *9*, 491–504.

Fujino, T., Goto, Y., Suga, H., and Murakami, H. (2013). Reevaluation of the d-Amino Acid Compatibility with the Elongation Event in Translation. J. Am. Chem. Soc. *135*, 1830–1837.

Goto, Y., Katoh, T., and Suga, H. (2011). Flexizymes for genetic code reprogramming. Nat. Protoc. *6*, 779–790.

Hecht, S.M., Alford, B.L., Kuroda, Y., and Kitano, S. (1978). "Chemical aminoacylation" of tRNA's. J. Biol. Chem. *253*, 4517–4520.

Hentzen, D., Mandel, P., and Garel, J.-P. (1972). Relation between aminoacyl-tRNA stability and the fixed amino acid. Biochim. Biophys. Acta - Nucleic Acids Protein Synth. *281*, 228–232.

Jakubowski, H., and Goldman, E. (1984). Quantities of individual aminoacyl-tRNA families and their turnover in Escherichia coli. J. Bacteriol. *158*, 769–776.

Lee, N., Bessho, Y., Wei, K., Szostak, J.W., and Suga, H. (2000). Ribozyme-catalyzed tRNA aminoacylation. Nat. Struct. Biol. *7*, 28–33.

Lohse, P.A., and Szostak, J.W. (1996). Ribozyme-catalysed amino-acid transfer reactions. Nature *381*, 442–444.

Louie, A., Ribeiro, N.S., Reid, B.R., and Jurnak, F. (1984). Relative affinities of all Escherichia coli aminoacyl-tRNAs for elongation factor Tu-GTP. J. Biol. Chem. *259*, 5010–5016.

Murakami, H., Ohta, A., Ashigai, H., and Suga, H. (2006). A highly flexible tRNA acylation method for non-natural polypeptide synthesis. Nat. Methods *3*, 357–359.

Pedersen, S., Bloch, P.L., Reeh, S., and Neidhardt, F.C. (1978). Patterns of protein synthesis in E. coli: a catalog of the amount of 140 individual proteins at different growth rates. Cell *14*, 179–190.

Ravel, J.M., Wang, S.-F., Heinemeyer, C., and Shive, W. (1965). Glutamyl and Glutaminyl Ribonucleic Acid Synthetases of Escherichia coli W : SEPARATION, PROPERTIES, AND STIMULATION OF ADENOSINE TRIPHOSPHATE-PYROPHOSPHATE EXCHANGE BY ACCEPTOR RIBONUCLEIC ACID. J. Biol. Chem. *240*, 432–438.

Roy, S. (1983). A continuous spectrophotometric assay for Escherichia coli alanyl-transfer RNA synthetase. Anal. Biochem. *133*, 292–295.

Saito, H., Watanabe, K., and Suga, H. (2001). Concurrent molecular recognition of the amino acid and tRNA by a ribozyme. RNA *7*, 1867–1878.

Shimizu, Y., and Ueda, T. (2010). PURE Technology. In Cell-Free Protein Production SE - 2, Y. Endo, K. Takai, and T. Ueda, eds. (Humana Press), pp. 11–21.

Smolskaya, S., Zhang, Z.J., and Alfonta, L. (2013). Enhanced Yield of Recombinant Proteins with Site-Specifically Incorporated Unnatural Amino Acids Using a Cell-Free Expression System. PLoS One *8*, e68363.

Soutourina, J., Plateau, P., and Blanquet, S. (2000). Metabolism of D-aminoacyl-tRNAs in Escherichia coli and Saccharomyces cerevisiae cells. J. Biol. Chem. *275*, 32535–32542.

Sprinzl, M. (2006). Chemistry of aminoacylation and peptide bond formation on the 3′terminus of tRNA. J. Biosci. *31*, 489–496.

Wang, H.H., Huang, P.-Y., Xu, G., Haas, W., Marblestone, A., Li, J., Gygi, S.P., Forster, A.C., Jewett, M.C., and Church, G.M. (2012). Multiplexed in vivo His-tagging of enzyme pathways for in vitro single-pot multienzyme catalysis. ACS Synth. Biol. *1*, 43–52.

Wu, M.X., and Hill, K.A.W. (1993). A Continuous Spectrophotometric Assay for the Aminoacylation of Transfer RNA by Alanyl-Transfer RNA Synthetase. Anal. Biochem. *211*, 320–323.

Yamane, T., Miller, D.L., and Hopfield, J.J. (1981). Discrimination between D- and L-tyrosyl transfer ribonucleic acids in peptide chain elongation. Biochemistry *20*, 7059–7064.

Chapter 3: Ribosomal engineering - Mutagenesis and Selection

Summary

In this chapter, we describe a novel method to derivatize ribosome 23S rRNA into a library containing defined rate of insertion, deletion and randomization mutations. We also designed and implement a 12-round selection scheme with gradual elevating selection pressure on both the bulkiness of amino acid D-side chain and the incorporation frequency of these amino acids on synthesized peptide. We found that the selection led to pool of mutants with activities diminished comparing to wild type but even comparing to each other on the synthesis of ^{D}Phe containing peptide. Future engineering to restore the entropic effect of peptidyl transfer catalysis would be needed to generate ribosome capable to efficiently incorporate D-amino acid.

Introduction

Previously we demonstrated two new approaches, i.e. optimizing aa-tRNA:EF-Tu interaction and relaxing EF-Tu aa-binding pocket, to improve the incorporation of D-aa into protein translation. In this chapter we describe the third attempt: engineering ribosome. Dedkova et al. has pioneered in this field by showing that mutations in core peptidyl transfer center (PTC, 2447-2450 and 2457-2462) could result in increased amber codon suppression corresponding to D-aa-tRNA (Dedkova et al., 2003). As a follow-up, we aim to build a platform to discover potential novel mutants with higher D-aa incorporation capacity in a high-throughput fashion. In protein engineering, two principles are often exercised, one is to translate the selection pressure into viability or binding affinity in order to enrich the desired mutants in batch from a population, and the second is to connect phenotype to genotype, so that active mutants could be identified conveniently by sequencing their genotypes. Since, the peptidyl transfer center of ribosome is all composed by 23S rRNA (domain V), a library of mutagenized rRNA sequences play both

phenotypes and genotypes, and hence the 2nd criterion is fulfilled by default. In order to turn D-aa incorporation activity into a binding assay, we constructed a ribosome display platform based on two facts: a) ribosomes stall as stable complex with mRNA and nascent peptide when it cannot incorporate more amino acids or terminate properly; b) the nascent peptide needs to grow to ca. 40 residues to have the first residue been accessible (Forster et al., 2004). The scheme in Figure 11 illustrates our platform. The N-terminus FLAG-tag serves as handle to pull down active mutants, but it would only be exposed when ribosome read-throughs a few codons reassigned to D-aas (or Uaa in general). The enriched mutant pool could then be recycled to the next round of selection and sequenced.

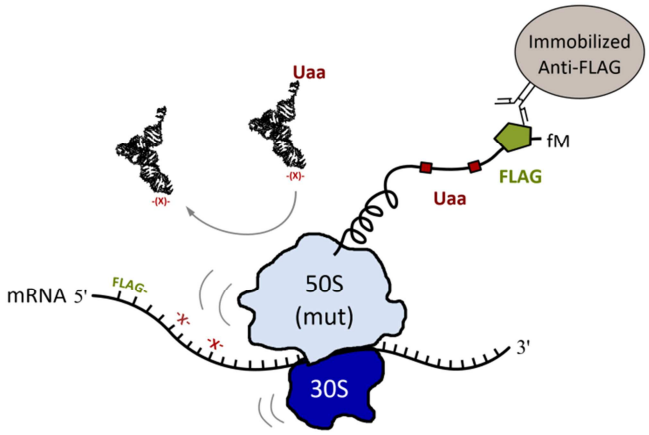


Figure 11. Selection platform for ribosome mutants based on ribosome display. The N-terminus FLAG-tag serves as handle to pull down active mutants, which will only expose when ribosome read-through a few codon reassigned to U-aas, including Daa and α, α-dialkylamino acid.

Results

Design and construction of library

While choosing regions of rRNA sequence to mutate, we notice that almost all positions proximal or proven critical to peptide transfer center (Fernández-Suárez et al., 2014) of 23S are highly conserved (Cannone et al., 2002), which implies traditional saturated mutagenesis strategy

would easily results in double or triple whammy in activity. From the recent crystal structure (Voorhees et al., 2009), we also notice that in several regions around peptide transfer center, rRNA is packed as loop or coil instead of complementary stem, so changing the crowdedness might impact as well as swapping specific inter-base hydrogen bonding. Therefore, utilizing recent technology advance in massive parallel oligo synthesis, we wrote script to render all mutated sequences with assigned mutation rate of each type, such as insertion, deletion, and base change, had one million species of them be synthesized on one microchip, and assembled to full-length 23S rRNA gene. The rRNA residues within 20 Å distance to PTC are picked and subjected to mutating scripts no matter what their local secondary structures are. The list of regions containing mutations is shown in Table 6 below. Mutation rates are set as: deletion 0.05%, insertion 0.03%, and randomization 0.015% for all regions.

Region	Num. of residues	Region	Num. of residues	Region	Num. of residues
U566-G577	12	A1668-A1679	12	G2271-C2275	5
G674-C678	5	U1779-A1786	8	G2436-A2461	26
G742-U755	14	A1937-A1966	30	G2490-U2511	22
G777-A793	17	G1983-C1985	3	C2551-C2558	8
U827-G830	4	G2027-G2035	9	G2569-C2611	43
G956-U963	8	A2052-C2073	22		
C1251-U1255	5	A2247-G2256	10	Total	263

Table 6. List of regions and numbers of residues selected for mutation.

With pools of synthetic oligos containing numerous mutations, next we consider the construction of ribosome library. The mutagenesis of prokaryote ribosome in vivo has been greatly limited by its indispensable role in massive and accurate protein production for cell survival; however, our attempts to assemble ribosome *in vitro* from *in vitro* prepared rRNA and ribosomal protein parts also suffer from very low efficiency possibly due to multiple rRNA base modifications. Given that our goal to mutate the core 23S rRNA will inevitably compromise ribosome activity in standard peptide bond synthesis, if we prepare ribosome library *in vitro*

would likely lead to a double whammy. We therefore use a transient expression method (Youngman and Green, 2005) to express the ribosome library *in vivo*, and inducing mutants with a titratable tetO-promoter to ca. 10% of total ribosome population, and followed by isolation mutants from constituent ribosome through the MS2-tag inserted on loops of 23S rRNA (Figure 12).

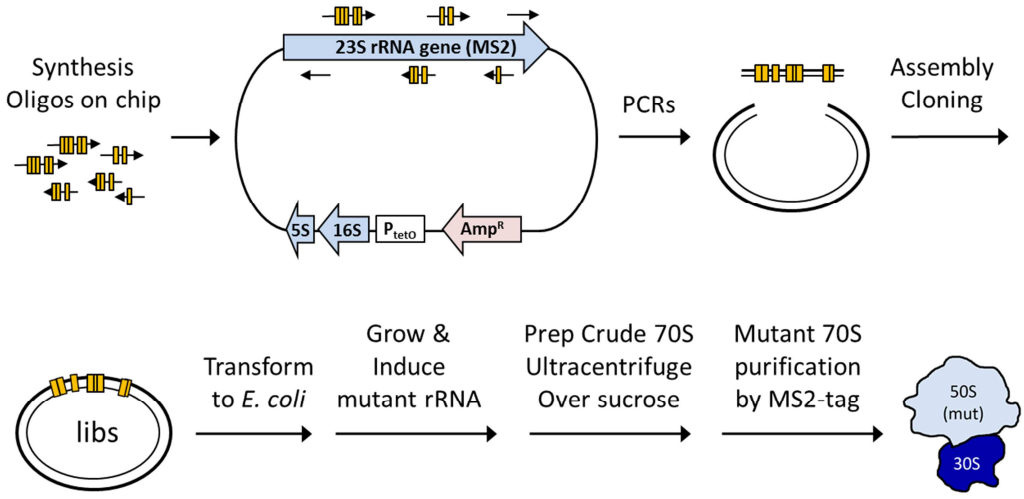
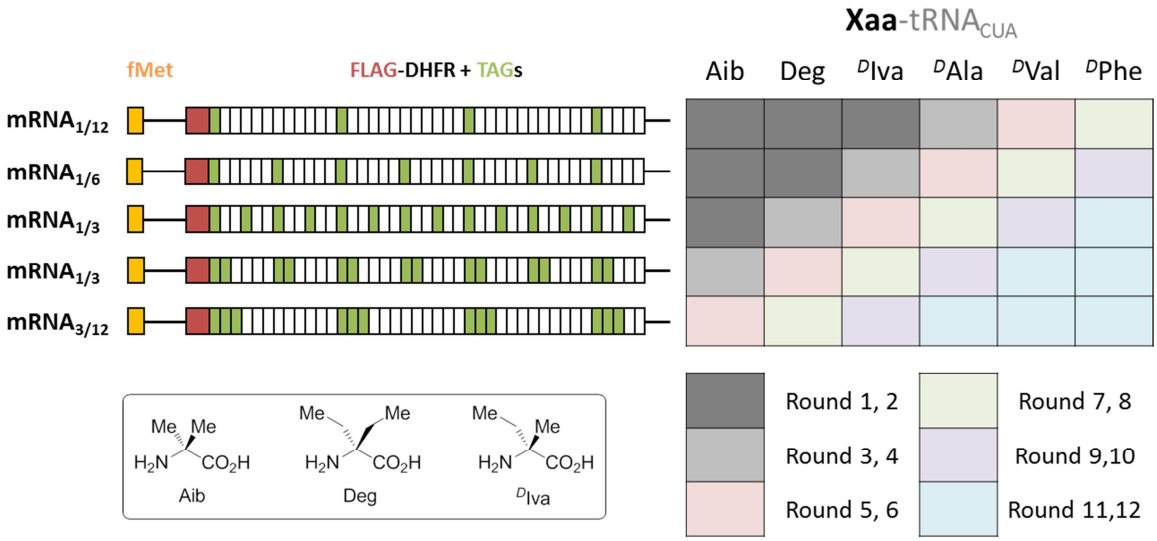


Figure 12. Illustration of our library preparation procedure. Chip oligos are joined by overlape-extention PCR into one 2 kb piece, and cloned into plasmid with the rest of rRNA genes from *rrnB* operon under *tetO* promoter. Library of plasmids are transformed in batch, and purified in two steps to yield mutant 70S ribosome.

Selections with stepwise increasing stringency

Our *in vitro* ribosome selection is based on the amber-codon suppression by unnatural amino acid in the expression of peptides beginning with a FLAG-tag, as summarized in Figure 11. The amber codons are placed in locations closely after FLAG-tag coding region so that mutants capable to incorporate unnatural substrate will read though and hence able to display FLAG epitope from the nascent peptide channel. At once, we attempted to use D-Lysine(Biotin) as handle instead of FLAG-tag, so if mutants can incorporate it into peptide, the ribosome-peptide-mRNA complex can be pull down. The method was later replaced by FLAG-tag

because of its low efficiency and inflexibility. But in that experiment, we found that we always get wild type sequence dominated in all tested condition. To explain, since in this experiment, at least 40 subsequent L-aa-to-L-aa peptide pond formations are required to have the handle exposed from ribosome exit tunnel; this gave ribosomes 40:1 selection pressures on L- versus D-peptide bond formation. While adding more amber codons in the template could help to balance L versus D-peptide bond amount, the low efficiency of D-amino acid incorporation might cause trouble. Therefore, we designed and carried through a selection consists of 12 rounds, in which the ribosome libraries translated templates round by round with increasing frequency of single amber codon, i.e. from 1 over every 12 codons to 1 over every 3 codons, increasing frequency of successive amber codon, and with increasing chirality of amino acids, i.e. from 2, 2-dimethyl to D-2-ethyl-2-methyl-amino acid, D-Ala, D-Val, and finally to D-Phe[4]. This is illustrated in Figure 13 below.



[4] We thought D-Phe is one of the difficult-to-incorporate D-aa when we performed this experiment. Later report from Fujino et al. shows D-Phe was actually relatively easy-to-incorporate among all 19 D-aas (Fujino et al., 2013).

Figure 13. Selection scheme. On the left is the five different templates used, the label indicates the frequency of amber codon and the number of consecutive amber codons. One the right is the table of amino acids and templates used in the 12 rounds of ribosome displays.

After every two cycles, rRNA of the pool of selected mutants is extracted and analyzed.

Sequence analysis and functional assay

We analyzed the 23S rRNA of mutants pull-down by anti-FLAG antibody by Sanger sequencing, and the interesting mutants are cloned individually for further characterization. We once attempted to adapt two next-generation sequencing methods, 454 pyrosequencing and PacBio's SMRT sequencing into our analysis but unfortunately both of their error rates (10% in 454, and 3% in PacBio) are too high to cope with. After we carried out 12 rounds of selection and enrichment procedure, around 90 sequences acquired from every two rounds are aligned by clustalW2 algorithm and represented as phylogenetic trees (Letunic and Bork, 2011) (Figure 14). The first observation among these trees is that a significant and distant clade emerged on round 6 and reappears in all later rounds. When merging sequences from round 6 to 12 into a new tree, these distant clades can all traced back to the same ancestor. Along the selection cycles, if disregarding those outlier clades, the overall distances among species reduced, with the extremely distant species removed. Further examination of the outlier clades indicates that they contains much more mutations on regions not covered by chip oligos, and the Blast of their sequences against NCBI nucleotide database shows that they are 99% identical to *Pseudomonas aeruginosa* 23S rRNA.

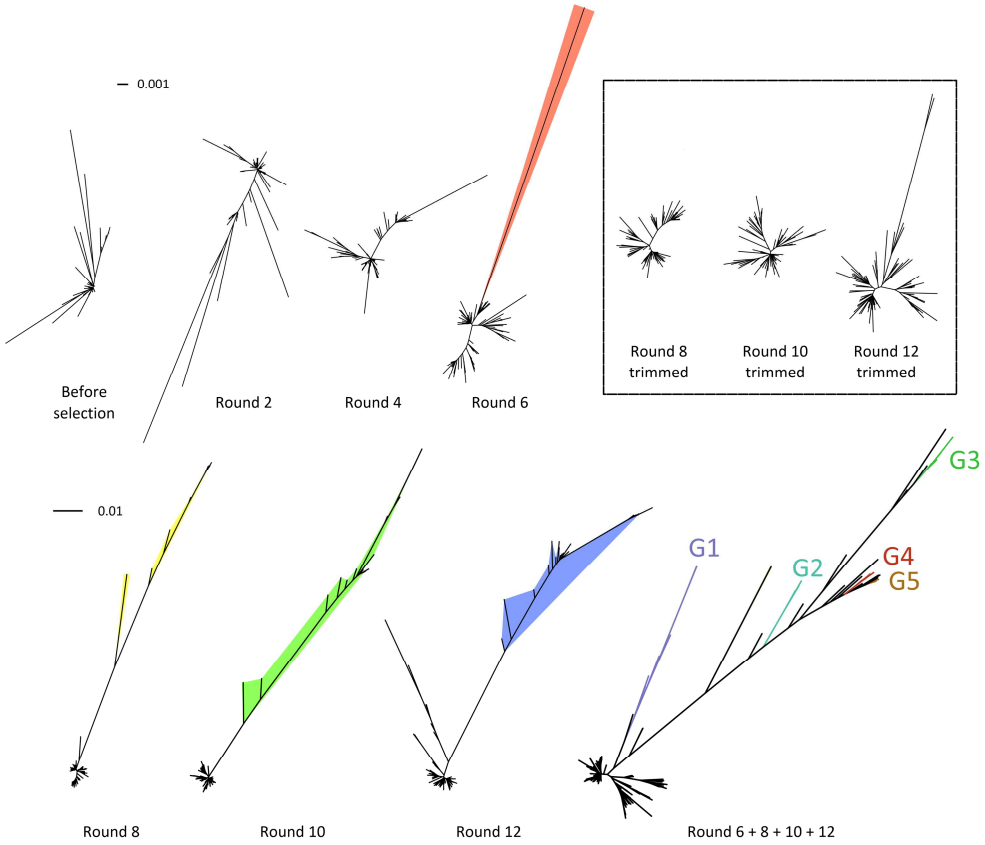


Figure 14. Phylogenetic analysis of sequence data collected from every two rounds of selection. Outlier clades appear since round 6 are colored. Tree plot without outlier clades for round 8, 10, 12 are showed in box. Outlier mutants picked are shown in trees from merged pool of round 6, 8, 10 and 12. Sequence of G3 are found to be 99% identical to *Pseudomonas aeruginosa* 23S rRNA, while G1, 2, 4 and 5 are various "DNA shuffling" hybrids between *P. aeruginosa* and *E. coli* 23S rRNA sequence. Mutant NP1 to NP5 are randomly picked from tree of round 12 without *P. aeruginosa* clade.

We arbitrarily pick five sequences from the merged tree (named as G1 to G5, taken from clades with color background, Figure 14), and another five spequences (NP1 to NP5, as for non-*Pseudomonas*) from the round 12 tree with contaminating *P. aeruginosa* removed. Their aligned sequences and list of mutations are presented in Appendix C. In brief, mutants G1 to G5 which have various degree of DNA sequence shuffling from *P. aeruginosa* contain 108-275 mutations with the majority on the surface of 50S ribosome subunit. On the other hand, mutants NP1 to NP5 contains only 8-24 mutations, but still, the majority of them are not result from originally

designed oligos. We prepared the corresponding mutant ribosomes and test their capacity on translation and D-aa incorporation.

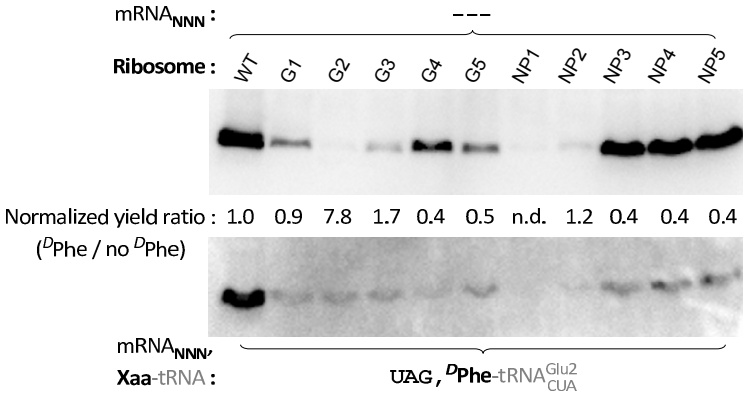


Figure 15. Assay for translation activity and ^{D}Phe incorporation of several random picked mutants. *P. aeruginosa* related mutants G1-G5 show various translation activity on normal peptides, and low but even activities on peptides containing ^{D}Phe. Non-*P. aeruginosa* related mutants also have various translation activities, with NP2 shows slightly higher yield ratio toward ^{D}Phe containing peptide. mRNA with label "---" is a template without middle NNN codon and hence full-length peptide synthesis doesn't require chemically acylated tRNA.

As shown in Figure 15, to our surprise, the *P. aeruginosa* and *E. coli* hybrid ribosome mutants G1 to G5 shows various level of translation activity on peptide containing normal amino acids, and also shows lower but even overall efficiency when synthesizing peptide containing ^{D}Phe. Densitometric analysis is done by first measure the intensity ratio of bands corresponding to normal peptide and ^{D}Phe-containing peptide for every ribosome mutants, and then these ratios are normalized to wild-type ribosome (as 1.0). For non-*Pseudomonas* related mutants, only NP2 shows slightly higher yield ratio toward ^{D}Phe containing peptide comparing to wild-type, none of the other mutants performs better than wild-type. We are testing these mutants with more D-amino acids now to see if any of them are capable to translate more efficiently than wild-type ribosome.

Discussion

In this chapter, we aimed to explore the possibility of expanding ribosome's substrate tolerance, by generating a mutant library of 23S rRNA. On the selection platform, we used various amino acids including α,α-dialkylamino acids to gradually increase selection pressure. The additional advantage of using α,α-dialkylamino acid is that, since ribosome is more tolerable to them than to D-aa, this would enable us to use templates encoding more U-aa, i.e. more balanced templates in terms of the ratio of encoded U-aa versus L-aa.

We identified several interesting 23S rRNA mutants after the selection. First, the appearance of *Pseudomonas*-derived mutant is unexpected, and remains elusive, but it is surprising that swapping ca. 2000 bp out of 2953 bp 23S rRNA sequence to *Pseudomonas aeruginosa* in mutant G3 still left us ribosome with measurable activities. It has been shown that ribosome from *Pseudomonas sp.* can utilize *E. coli* translation factors and aa-tRNAs in cell extract based system (Landau et al., 1977), but swapping rRNA is unprecedented. Regarding mutants NP1 to NP5, it is also unclear to us how those non-designated mutations are generated; presumably they are errors from PCR amplification. For both types of mutants, although various activities are observed for expression regular peptide, they all show lower than wild-type activities in making ^{D}Phe-containing peptide, suggesting that mutations diminish ribosome's catalytic power.

Considering the substrate-assisted catalysis hypothesis and the *ab initio* modeling of the transition state of peptidyl transfer (Wallin and Åqvist, 2010; Weinger et al., 2004), catalysis of peptidyl transfer reaction is conferred mainly by entropic effect, and to less extent by proximal effect. Specifically, ribosome promotes peptide bond formation mainly by organizing water molecules, peptidyl-tRNA and aa-tRNA into proper conformations. Therefore, perturbation on this reaction either by designated mutations around PTC or by the use of D-aa-tRNA would

cause the loss of pre-organized conformation and hence the vast majority of catalytic power. That is to say, replacing L-aa by D-aa not only creates steric clash in PTC pocket but the misaligned hydrogen bonding which greatly deteriorates the entropic effect. This could explain why some mutants, such as G3 and G5, express normal peptide as worse as peptide containing D-aa, as they probably already lose entropic catalysis effect.

Therefore, to re-build entropic effect-based catalysis might be one important direction to consider. Current wild-type ribosome architecture is well-suited for L-aa peptidyl transfer but stereo specificity is contributed to a great extent by L-aa-tRNAs themselves. To build a ribosome that catalyzes both D-aa and L-aa efficiently would likely involve a different mechanism, or potentially require help from other protein factors, such as the help of elongation factor P on translation of consecutive proline residues (Doerfel et al., 2013; Ude et al., 2013).

Materials and Methods

Construction of rRNA plasmid library

Plasmid p278MS2 contains entire *rrnB* operon under promoter P_L and with 23S rRNA tagged by MS2 hairpin is kindly provided by R. Green (Youngman and Green, 2005). A new plasmid ptet278 is constructed by first digesting p278MS2 with MfeI and KpnI, and swapping its promoter to tetO by reassembly with dsDNA oligos: 5'-AATTCAATTG TCCCTATCAG TGATAGAGTT GACATCCCTA TCAGTGATAG AGATAATGGG TACCGTAC-3', using GeneArt® Seamless Cloning kit (Life Technology). The ptet278 plasmid is amplified with primer RID-11T and RID-11B (Table 8) to acquire the plasmid backbone except ca. 2 kb region of 23S gene to be mutated.

Mutagenesis oligos pools are synthesized by Custom Array® 12k format, with sequences generated by python script (Appendix A) based on sequences from eight 170 bp regions of 23S rRNA (Table 7). Mutation rate are set as below: insertion 3%; deletion 5%, randomization 1.5%.

Oligo	Sequence
RID-1	TAGGCGTGTGACTGCGTACCTTTTGTATAATGGGTCAGCGACTTATATTCTGTAGCAAGGTTAAC CGAATAGGGGAGCCGAAGGGAAACCGAGTCTTAACTGGGCGTTAAGTTGCAGGGTATAGACCC GAAACCCGGTGATCTAGCCATGGGCAGGTTGAAGGTTGGGTA
RID-2	ACACTAACTGGAGGACCGAACCGACTAATGTTGAAAAATTAGCGGATGACTTGTGGCTGGGGGTT GAAAGGCCAATCAAACCGGGAGATAGCTGGTTCTCCCCGAAAGCTATTTAGGTAGCGCCTCGTG AATTCATCTCCGGGGGTAGAGCACTGTTTCGGCAAGGGGGTC
RID-3	ATCCCGACTTACCAACCCGATGCAAACTGCGAATACCGGAGAATGTTATCACGGGAGACACACG GCGGGTGCTAACGTCCGTCGTGAAGAGGGAAACAACCCAGACCGCCAGCTAAGGTCCCAAAGT CATGGTTAAGTGGGAAACGATGTGGGAAGGCCCAGACAGCCAG
RID-5	TGAGAGAACTCGGGTGAAGGAACTAGGCAAAATGGTGCCGTAACTTCGGGAGAAGGCACGCTG ATATGTAGGTGAGGTCCCTCGCGGATGGAGCTGAAATCAGTCGAAGATACCAGCTGGCTGCAAC TGTTTATTAAAAACACAGCACTGTGCAAACACGAAAGTGGACG
RID-7	GGTCCTAAGGTAGCGAAATTCCTTGTCGGGTAAGTTCCGACCTGCACGAATGGCGTAATGATGG CCAGGCTGTCTCCACCCGAGACTCAGTGAAATTGAACTCGCTGTGAAGATGCAGTGTACCCGCG GCAAGACGGAAAGACCCCGTGAACCTTTACTATAGCTTGACA
RID-8	AGGCTTTGAAGTGTGGACGCCAGTCTGCATGGAGCCGACCTTGAAATACCACCCTTTAATGTTTG ATGTTCTAACGTTGACCCGTAATCCGGGTTGCGGACAGTGTCTGGTGGGTAGTTTGACTGGGGCG GTCTCCTCCTAAAGAGTAACGGAGGAGCACGAAGGTTGGC
RID-9	CGCGAGCAGGTGCGAAAGCAGGTCATAGTGATCCGGTGGTTCTGAATGGAAGGGCCATCGCTCA ACGGATAAAAGGTACTCCGGGGATAACAGGCTGATACCGCCCAAGAGTTCATATCGACGGCGGT GTTTGGCACCTCGATGTCGGCTCATCACATCCTGGGGCTGAA
RID-10	GTAGGTCCCAAGGGTATGGCTGTTCGCCATTTAAAGTGGTACGCGAGCTGGGTTTAGAACGTCG TGAGACAGTTCGGTCCCTATCTGCCGTGGGCGCTGGAGAACTGAGGGGGGCTGCTCCTAGTACG AGAGGACCGGAGTGGACGCATCACTGGTGTTCGGGTTGTCAT

Table 7. Sequences used to generate rRNA mutagenesis oligos. Underscore bases are regions to be mutated. Fragment RID-4 and RID-6 contains no residue to mutate.

Primer	Sequence
RID-1T	AGCACGCTTAGGCGTGTGACTGCGTACC
RID-1B	CTCCAGTTAGTGTTACCCAACCTTCAACCTGCC
RID-2T	TTGAAGGTTGGGTAACACTAACTGGAGGACCGAA
RID-2B	GTAAGTCGGGATGACCCCCTTGCCGAAACAGT
RID-3T	AAGGGGGTCATCCCGACTTACCAACCCGAT
RID-3B	CCAACATCCTGGCTGTCTGGGCCTTCCC
RID-4T	AGGCCCAGACAGCCAGGATGTTGGCTTA
RID-4B	CACCCGAGTTCTCTCAAGCGCCTTGGTATTC
RID-5T	GGCGCTTGAGAGAACTCGGGTGAAGG
RID-5B	ACACCGTATACGTCCACTTTCGTGTTTGCACAG
RID-6B-2	TTAGGACCGTTATAGTTACGGCCGC
RID-6B	TTAGGACCGTTATAGTTACGGCCGCCGTTTACCGGGGCTTCGATCAAGAGCTTCGCTTGCGCTAA CCCCATCAATTAACCTTCCGGCACCGGGCAGGCGTCACACCGTATACGTCCACTTTCGTG
RID-7T	CGGCCGTAACTATAACGGTCCTAAGGTAGCG
RID-7B	CCTACACATCAAGGCTCAATGTTCAGTGTCAAGCTATAGTAAA
RID-8T	CTGAACATTGAGCCTTGATGTGTAGGATAGGTGGGAGGCTTTGAAGTGTGGACGC

RID-8B	TATGCCATTGCACTAACCTCCTGATGTCCGACCAGGATTAGCCAACCTTCGTGCTCCTCC
RID-9T	GACGGCGCGAGCAGGTGCGAAAG
RID-9B	TGGGACCTACTTCAGCCCCAGGATGTGATG
RID-9T-2	GAGGTTAGTGCAATGGCATAAGCCAGCTTGACTGCGAGCGTGACGGCGCGAGCAGGTGCG
RID-10T	TGGGGCTGAAGTAGGTCCCAAGGGTATGG
RID-10B	CCATTGGCATGACAACCCGAACACCAGT
RID-11T	GTGTTCGGGTTGTCATGCCAATGGCACTGCCCGGTAGC
RID-11B	CGCAGTCACACGCCTAAGCGTGCTCCCACTGCTTGTACGT
rRT1	AGGACCCTACTAGTAAAGATGGGTAATCCTCATCA

Table 8. Primers used to amplify rRNA mutagenesis oligos and assemble them into one fragment.

From RID-1 to 10, each fragment are amplified by corresponding primers with KAPA™ SYBR® FAST qPCR Kits till mid-log phase. Fragment RID-4 and RID-6 contains no mutation, so plasmid ptet278 is used as template. They are assembled by overlap-extension PCR and clone to linearlized ptet278 plasmid to yield plasmid library in DH5α-pro which has high level of tetR to suppress mutant rRNA expression.

Purification of ribosome library or individual mutant

Plasmid pMS2-MBP-His is kindly provided by R. Green's group, and protein MS2-MBP is overexpressed and purified according to their procedure (Youngman and Green, 2005).

Mutant library or individuals are grown in 1 L of 2YTPG media (1.6% Yeast Extract, 1% Tryptone, 0.5% NaCl, 5 g/L glucose, 22.05mM KH_2PO_4 and 39.26 mM K_2HPO_4) with 100 μg/mL Carbenicillin in 2.5 L Tunair® flask and harvested at OD_{600} ~ 3.0. The wet cell paste was resuspend in buffer A (20 mM Hepes-KOH, pH 7.6, 100 mM NH_4Cl, 10 mM $Mg(OAc)_2$, and 0.5 mM EDTA) and pass through a French Press, and cell debris was removed by centrifugation twice at 30,000g for 20 min. Supernatants are layered over 10 mL buffer B (20 mM Hepes-KOH, pH 7.6, 500 mM NH_4Cl, 10 mM $Mg(OAc)_2$, 0.5 mM EDTA and 37.7% Sucrose) and spin at 45,000 rpm in rotor Type 70 Ti for 12 hr. Crude ribosome pellet is resuspended in buffer A without EDTA, loaded over 1 mL amylose column pre-bound with MS2-MBP protein and

washed with 10 mL buffer A without EDTA. Mutant ribosomes are eluted with buffer A without EDTA plus 10 mM Maltose, concentrated, and dialysis against storage buffer (20 mM Hepes-KOH, 30mM HCl, 10 mM $Mg(OAc)_2$ and 7 mM 2-βME).

Selection of mutants by ribosome display

Translations are done with customized PURExpress kit (New England Biolab Inc.), in which release factors, ribosome, amino acids and tRNAs are dropped out from other components. DNA templates sequences are listed in Appendix B. A typical 50 μL translation reaction contains: 1x Solution A, 1x Factor mix, 0.1 mM each amino acids, 1.85 mg/mL total tRNA (Roche), 1μM ribosome library, 2 μM photo-deprotected aa-tRNA and 10 ng/μL DNA template. The mixture is incubated at 37 °C for 1 hr, shacked in anti-FLAG microtiter plate for 30 min, 750 rpm at 30 °C, and washed with WBT buffer (50 mM Tris-HOAc pH 7.4, 150 mM NaCl, 50 mM $Mg(OAc)_2$, 0.05% Tween-20) four times. Bound ternary complex is dissociate with 100 μL denature buffer (300 mM NaOAc pH 5.5, 0.25% SDS, 15 mM EDTA, 50 μg/mL Yeast tRNA) and denatured by vortexing with 100 μL acid phenol for 2 min. The aqueous phase is washed once more with phenol and twice with 24:1 chloroform-isoamyl alcohol followed by ethanol precipitation. Resuspended rRNA is reverse-transcribed by SuperScript® III reverse transcriptase (Life Technology) with primer rRT1 and amplified by primer RID-1T and RID-10B. This mixture of fragment is then clone back to linearized ptet278 backbone and transformed to DH5α-pro for next round of selection of sequencing analysis.

Sequence analysis

Sequence of each mutant is assembled from three Sanger sequencing reads. Sequences from every 2 rounds of selection are pooled and aligned by Muscle Command line algorithm.

The alignment files are loaded onto the ClustalW2-Phylogeny server on EMBL-EBI website to generate phylogenetic trees. Trees are edited by TreeGraph 2 (Stöver and Müller, 2010) and plotted by interactive Tree of Life webserver (Letunic and Bork, 2011).

Western blot and translation activity assay

Same procedure as described in chapter 2 is followed.

References

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L. V, Müller, K.M., et al. (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics *3*, 2.

Dedkova, L.M., Fahmi, N.E., Golovine, S.Y., and Hecht, S.M. (2003). Enhanced d-Amino Acid Incorporation into Protein by Modified Ribosomes. J. Am. Chem. Soc. *125*, 6616–6617.

Doerfel, L.K., Wohlgemuth, I., Kothe, C., Peske, F., Urlaub, H., and Rodnina, M. V (2013). EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. Science *339*, 85–88.

Fernández-Suárez, X.M., Rigden, D.J., and Galperin, M.Y. (2014). The 2014 Nucleic Acids Research Database Issue and an updated NAR online Molecular Biology Database Collection. Nucleic Acids Res. *42* , D1–D6.

Forster, A.C., Cornish, V.W., and Blacklow, S.C. (2004). Pure translation display. Anal. Biochem. *333*, 358–364.

Fujino, T., Goto, Y., Suga, H., and Murakami, H. (2013). Reevaluation of the d-Amino Acid Compatibility with the Elongation Event in Translation. J. Am. Chem. Soc. *135*, 1830–1837.

Landau, J. V, Smith, W.P., and Pope, D.H. (1977). Role of the 30S ribosomal subunit, initiation factors, and specific ion concentration in barotolerant protein synthesis in Pseudomonas bathycetes. J. Bacteriol. *130*, 154–159.

Letunic, I., and Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. *39*, W475–W478.

Stöver, B.C., and Müller, K.F. (2010). TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics *11*, 7.

Ude, S., Lassak, J., Starosta, A.L., Kraxenberger, T., Wilson, D.N., and Jung, K. (2013). Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. Science *339*, 82–85.

Voorhees, R.M., Weixlbaumer, A., Loakes, D., Kelley, A.C., and Ramakrishnan, V. (2009). Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. Nat. Struct. Mol. Biol. *16*, 528–533.

Wallin, G., and Åqvist, J. (2010). The transition state for peptide bond formation reveals the ribosome as a water trap. Proc. Natl. Acad. Sci. *107*, 1888–1893.

Weinger, J.S., Parnell, K.M., Dorner, S., Green, R., and Strobel, S.A. (2004). Substrate-assisted catalysis of peptide bond formation by the ribosome. Nat. Struct. Mol. Biol. *11*, 1101–1106.

Youngman, E.M., and Green, R. (2005). Affinity purification of in vivo-assembled ribosomes for in vitro biochemical analysis. Methods *36*, 305–312.

Chapter 4: Synthesis of 2'-deoxy-3'-mercapto-tRNA Substrate for Ribosomal Polyketide Synthesis

Summary

We fused polyketide substrate, i.e. malonyl thioester to tRNA 3'-terminus, in order to examine if we could realize template-directed polypeptide or polyketide synthesis by the *in vitro* translation system. Synthesis of 2'-deoxy-3'-S-malonyl-tRNA substrate is achieved, but the preliminary attempts to incorporate it into ribosomal peptide synthesis is not successful. The labile thioester bond is suspected to cause this issue, and hence several strategies have been proposed to address it.

Introduction

Many strategies have been developed in an attempt to make artificial polyketides, due to their high potential clinical value, such as combinatorial synthesis or biosynthesis (McDaniel et al., 1999; Menzella et al., 2005). Split-pool solid-phase combinatorial synthesis (Goess et al., 2006) was able to produce more than 10,000 new compounds, but since same chemistries are applied at each diversify step, most of the library compounds share similar skeletons. Genetic engineering of the polyketide synthesis (PKS) pathway, on the other hand, has a record of building over 50 variants of 6-deoxyerythronolide B (DEBS) (Xue et al., 1999). However, it is only in a few systems like DEBS that we know thorough information about the domain DNA sequences and their spatial orientations in the modules that lead to the consecutive processing, where we could then rationally engineer the pathway to have reasonable yields (Weissman and Leadlay, 2005).

We aim to develop a new strategy for building more diversified polyketide skeletons by integrating ketide and peptide synthesis, two types of molecule that span distant structural and

functional space, into a single, programmable platform, i.e. ribosomal translation. Protein translation and polyketide assembly-line synthesis share a similar polymer extension mechanism (Figure 11a, c) (Ferrer et al., 1999; Gindulyte et al., 2006), although one follows information from RNA templates and the other works based on modular enzyme architecture. Intrigued by the diverse reactivity and broad substrate categories of PKS, we attempted to transplant the core of polyketide synthesis unit and fused it into the integrated ribosome translation machineries to collect the merits of both. Specifically, we would like to swap the peptidyl transfer reaction from amide formation to Claisen condensation which is the typical polyketide extension reaction. The chemical acylation methods described in previous chapters are adapted to prepare tRNA malonyl substrates that mimic polyketide synthesis. In this chapter we will describe how tRNA with 3'-hydroxy or 3'-mercapto group is synthesized and acylated by malonyl ketide building block and our preliminary attempts on subjecting them into *in vitro* translation system.

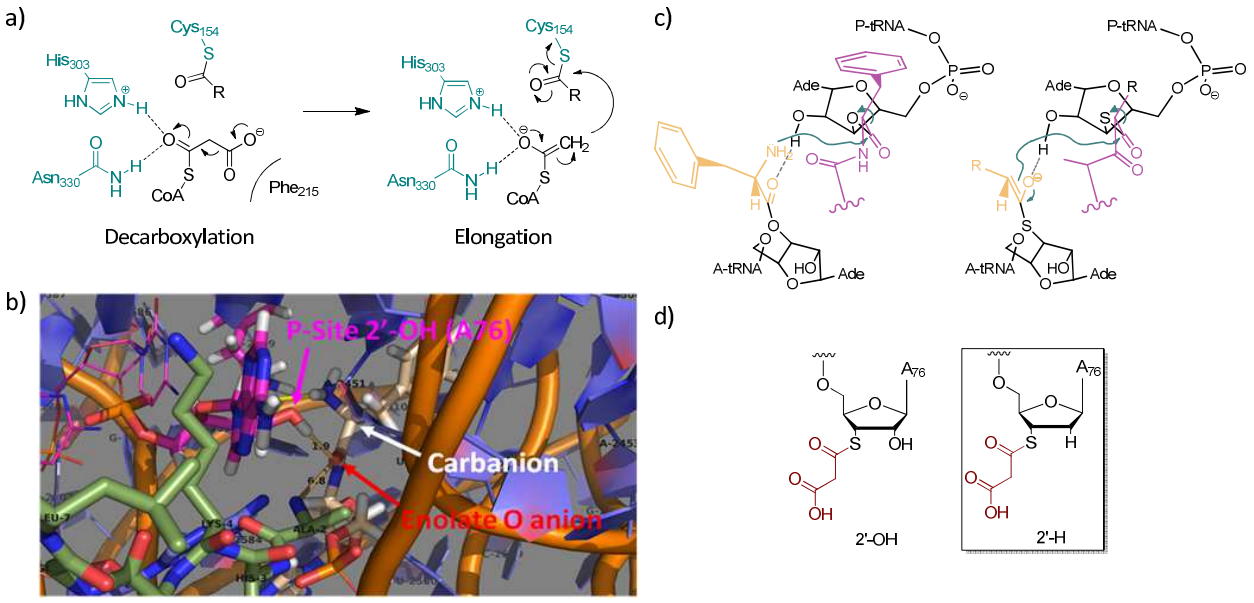


Figure 16. (a) mechanism of chalcone synthetase, redrawn from Ferrer et al. (Ferrer et al., 1999) (b) sketch representation of acyl transfer of peptide synthesis and its ketide synthesis mimic, drawn based on structure solved by Voorhees R. et al (Voorhees et al., 2009) (c) contexts of (b), magenta sticks: P-site tRNA (Phe); wheat sticks: A-site tRNA (Phe); blue cartoon: *T. thermophilus* 23S

rRNA; pale green cartoon: L27 protein (d) structures of designed ketide building block carrier tRNA (PDB ID: 2WDK, 2WDL).

Results

Design and Synthesis of 2'-deoxy-3'-mercapto-tRNA

Polyketide synthetases use a thioester to carry out polymerization due to its good reactivity toward unstable or weak nucleophiles such as enolates or the hydroxyl group. Based on the mechanism of ketosynthetase described by Ferrer et al.(Ferrer et al., 1999) (Figure 11a), a histidine and an asparagine are required as hydrogen bond donors to stabilize the enolate intermediate formed after decarboxylation of the malonate half thioester. Direct substitution of the P- and A-site amino acid with thioester and enolate in recent ribosome X-ray crystal structure solved by Ramakrishnan's group (Voorhees et al., 2009) is shown in Figure 11b (and Figure 11c for their interaction with surrounding residues). It is found that the closest ribosomal protein, L27, is 6.8Å away from the peptide transfer center, and the nearest hydrogen bond donor available for an A-site enolate is the P-site 2'-OH. Based on this analysis, tRNAs with two types of terminal "carrier module" are designed (Figure 11d), however, the carrier 2'-OH is not further pursued since 2'-OH will result in migration of the acyl group (Porcher et al., 2005).

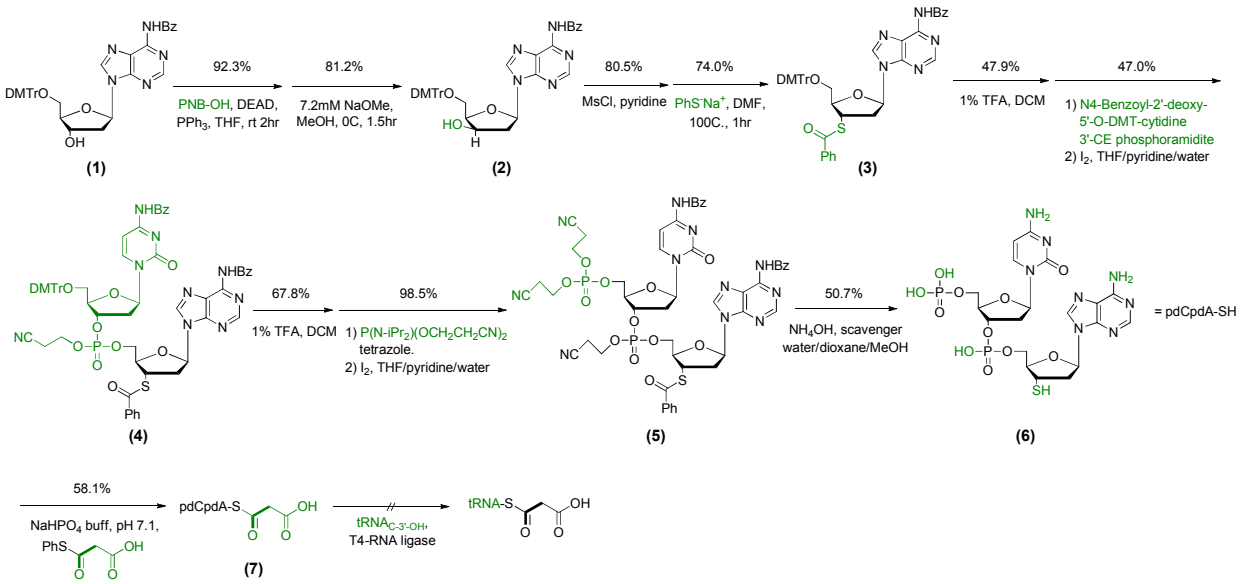


Figure 17. Synthesis scheme for the preparation of 2'-deoxy-3'-mercapto-tRNA.

PKS carrier tRNAs will be prepared by the 3'-terminus dinucleotide-ligation method first developed by Noren et al.(Ellman et al., 1991; Noren et al., 1989) The synthetic scheme of 3'-SH dinucleotides is shown in Figure 17. In brief, 3'-OH of commercial available 5'-DMT-adenosine **(1)** is first inverted by a Mitsunobu reaction, then mesylated and transformed to benzoyl-protected thiol **(3)** by S_N2 substitution. The 5'-DMT protection is removed and deoxycytidine as well as the terminal phosphate group are installed **(5)** via a standard oligonucleotide phosphoramidite method. After global deprotection with ammonium hydroxide, the resulting dinucleotides are loaded with malonate by aqueous transthioesterification **(7)**. However, the attempt to ligate pdCpdA-S-malonyl with *in vitro* transcribed tRNAs was not successful (Figure 18). In order to understand why the ligation fails, we prepared the disulfide dimer **(8)** of thiol **(6)** and also a *p*-hydroxyphenacyl (*p*HP) protected version of dinucleotide thiol **(9)**, and compare their effect on T4 RNA ligase activity.

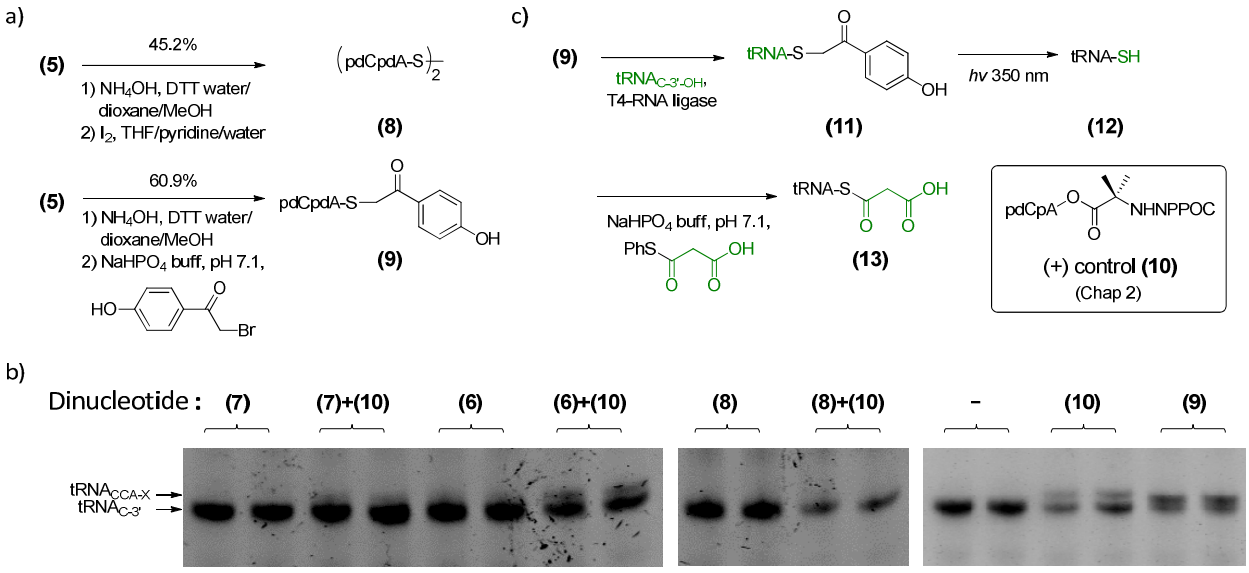


Figure 18. (a) Synthesis of disulfide dimer of dinucleotide pdCpdA-SH **(6)** and *p*-hydroxyphenacyl protected dinucleotide **(9)**; (b) 8% Acid Urea-PAGE analysis of dinucleotide ligation. pdCpdA-SH **(6)** is neither substrate nor inhibitor to T4 ligase; disulfide **(8)** and malonyl-S-pdCpdA **(7)** act as enzyme inhibitor, their presence block the ligation of normal substrate Aib-pdCpA **(10)**; *p*HP-masked thiol **(9)** is a good substrate. (c) Synthesis of 2'-deoxy-3'-mercapto-tRNA **(13)** from **(9)**.

On Figure 18c, when pdCpdA-S-malonyl substrate **(7)** is used, no ligation product, i.e. full length tRNA is detected, neither does pdCpdA-SH **(6)** substrate. However, an interesting effect is that, when mixing a standard dinucleotide substrate **(10)** into each ligation reaction, pdCpdA-S-malonyl **(7)** inhibits normal ligation of **(10)** to tRNA while pdCpA-SH **(6)** not. On the other hand, the disulfide dimer **(8)** also shows inhibitory effect on control ligation as pdCpdA-S-malonyl **(7)**. These suggest that T4 RNA ligase is sensitive to sulfhydryl modification reagents, either by oxidation or acylation, which is confirmed in literature (Eun, 1996). Although this couldn't explain why free thiol **(6)** is not recognized as substrate, it suggests that masking the thiol with inert group might work. Indeed, *p*HP-protected dinucleotide is ligated to tRNA reaches ca. 50 % yield. The protecting *p*HP group is removed by photolysis and the tRNA-3'-thiol is acylated by malonyl-half-thioester (MAHT) to build the 2'-deoxy-3'-mercapto-tRNA.

However, LC-MS data indicates malonyl-thioester of pdCpdA-SH **(7)** is very unstable, only 80% of **(7)** hydrolyzed to **(6)** within 5 min incubation in pH 7 buffer at room temperature. And no thioester **(7)** is detected 30s after mixing it with 5 mM 1, 4-dithioerythritol. When subject malonyl-S-tRNA **(13)** into the *in vitro* translation read-through assay as described in previous chapters, no signal is detected. We conclude that the following up experiment is to figure out a substitution for thiol reducing agents in PURE translation reaction, either biochemically by other reducing agents such as ascorbic acid, or physically degassing.

Discussion

A tool to incorporate polyketide units into protein synthesis and to integrate PKS decoration enzymes into the system will open tremendous new possibilities, not only that polyketides could be hybridized with peptides to make novel breeds of compounds but also that proteins could be dressed with ketides or vice versa for extended functional space. All of these could simply be achieved by programming the DNA templates. The accumulated studies on protein properties today would add great leverage to our ribosomal polyketides strategy in that one can add any "Functional Group" module to ketides to expand their functional span.

In this chapter, we described our attempt to integrate polyketide synthesis unit into ribosome translation system. A carrier tRNA with 3'-OH substituted by 3'-SH has been synthesized, and charged with standard polyketide substrate. However, since the thioester bond of tRNA-S-malonyl is labile in the presence of free thiol, further debugging is required. There are three intertwined issues to consider, one is to maintain reducing environment, since that scavenging DTT by adding Iodoacetamide seizes protein synthesis; another is the fast deacylation of thioester tRNA substrate needed to be addressed; and that high dosage of reactive

thioester such as tRNA-S-malonyl could potentially deactivate critical sulfhydryl group enzymes activity, ex. TrpRS. For the first issue, one idea would be to degas all PURE translation reaction thoroughly, but this would require meticulous operation. The other would be to find a non-thiol-based reducing agent that could protect enzyme activities while keeping dissolved oxygen level low, such as ascorbic acid. For the second and third issue, flexizyme might be a potential means to re-charge tRNA. The use of catalyst to recycle tRNA-SH not only reduces the initial concentration of tRNA-S-malonyl, but also allows to reload tRNA-SH with less reactive reagents such as *p*-chlorobenzyl mercaptan MAHT (instead of very reactive thiolphenol MAHT).

Materials and Methods

General

Solvents and reagents are purchased from Sigma Aldrich Inc. or EMD Merck without further purification. 1H NMR spectra were measured from Variant Mercury 400 MHz, Inova 500 MHz and Agilent DD2 600 MHz instrument. High Resolution Mass spectra (HRMS) were obtained from Agilent 6210 Time-of-Flight LC/MS. Preparative HPLC was carried out with an Agilent 1200 system equipped with diode array detector, fraction collector and Zorbax PrepHT C18 column (21.2 x 250 mm, 7 μm). Flash chromatography is done by CombiFlash® Rf 200 with either RediSep® Rf Gold normal phase silica or C18 column. Thin layer chromatography (TLC) is done with 5 × 2 cm glass plate coated by Silica Gel 60 with F_{254} fluorescence indicator.

Synthesis of compounds in this study

N^6-Benzoyl-5'-O-(4,4'-dimethoxytrityl)-2'-deoxyxyloadenosine (2): To a 250 mL three-necked flask were added N^6-Benzoyl-2'-deoxy-5'-O-DMT-adenosine (**1**) (5.0 g, 7.6 mmol,

1 eq.), *p*-nitrobenzoic acid (2.55 g, 15.2 mmol, 2 eq.), and triphenylphosphine (4.0 g, 15.2 mmol, 2 eq.) under nitrogen. Anhydrous THF (50 mL) was added to dissolve the reagents while stirring. After all material dissolved, diethyl azodicarboxylate (40% w/v in toluene, 6.9 mL, 15.2 mmol, 2 eq.) was added drop wise over five minutes. Reaction progress was followed by TLC in EtOAc. After stirred 2.5 hr, solvent was evaporated and the crude solid was purified by flash chromatography with EtOAc. Fractions corresponding R_f = 0.67 (EtOAc) are pooled and concentrated to yield *p*-nitrobenzoylated alcohol as a yellow solid (5.66 g, 7.01 mmol, 92.3%).

p-nitrobenzoylated alcohol (0.915 g, 1.34 mmol, 1 eq.) was then suspended in MeOH (130 mL) and cooled to 0 C, then a 0.2 M solution of NaOMe in MeOH (4.54 mL, 0.907 mmol, 0.8 eq.) was added drop wise over two minutes, which fully dissolves *p*-nitrobenzoylated alcohol. Reaction progress was monitored by TLC eluting with EtOAc. No starting material remained after 35 minutes, and the reaction was quenched by saturated NH_4Cl to pH 7 dropwisely and solvent was evaporated *in vacuuo* at room temperature yielding a crude pale yellow solid. The crude was purified by flash chromatographed on silica gel (EtOAc to 4:6 EtOAc/MeOH) to yield alcohol (**2)** as a transparent film (0.61 g, 0.93 mmol, 81.2%). R_f = 0.29 (EtOAc). ^{1}H NMR (400 MHz, DMSO) δ = 8.79 (s, 1H), 8.45 (s, 1H), 8.00-6.77 (m, 18H), 6.45 (d, 1H), 5.45 (d, 1H), 4.38 (m, 1H), 4.22 (m, 1H), 3.69 (m, 6H), 3.2 (d, 1H), 2.79 (m, 1H), 2.49 (d, 1H). HRMS calculated mass for $C_{38}H_{35}N_5O_6$: 657.2587; found $[M+H]^+$: 658.2758.

N^6-Benzoyl-5'-O-(4,4'-dimethoxytrityl)-3'-S-benzoyl-2',3'-dideoxyadenosine (3): Alcohol **(2)** (0.606 g, 0.921 mmol, 1 eq.) was dissolved in anhydrous pyridine (7.2 mL) under nitrogen and co-evaporate with 5 mL pyridine *in vacuuo* twice. The residue was then dissolved in anhydrous pyridine (8.5 mL) and cooled to 0 °C. To the solution was sequentially added dimethylaminopyridine (DMAP) (0.125 g, 1.01 mmol, 1.1 eq.) and mesyl chloride (78.4 μL,

1.013 mmol, 1.1 eq.) dropwisely over two minutes. The reaction was allowed to warm to room temperature and stirred overnight. The reaction mixture was pour into DCM and H_2O mixture and shacked vigorously. The organic phase is washed twice more with 0.1N HCl and dried with sodium sulfate, and then concentrated to yield pale orange crude solid (0.55 g), which was used without further purification. R_f = 0.52 (EtOAc). 1H NMR (600 MHz, $CDCl_3$) δ = 8.84 (m, 1H), 8.40 (s, 1H), 8.30 (s, 1H), 8.10-7.57 (m, 5H), 7.48-6.88 (m, 13H), 6.65 (m, 1H), 5.54 (s, 1H), 4.42 (m, 1H), 4.19 (m, 1H), 3.74 (m, 1H), 3.41 (m, 1H), 3.03 (m, 3H), 2.80 (s, 1H), 2.41 (s, 1H). HRMS calculated mass for $C_{39}H_{37}N_5O_8S$: 735.2363; found $[M+H]^+$: 736.2503.

In 250 mL three-neck flask, mesylated alcohol (2.53 g, 3.44 mmol) is co-evaporated with 10 mL anhydrous pyridine twice. 30 mL anhydrous DMF is added and the solution is heated under argon, and once the temperature reaches 100 °C, sodium thiobenzoate (1.37 g, 8.60 mmol) is added every 30 min for four times. 30 min after the last addition of sodium thiobenzoate, the reaction is followed by TLC and then cooled down to ca. 60 °C. DMF is slowly evaporated *in vacuuo*. The residue is dissolved in 250mL DCM and washed with 200mL of brine three times, dried over Na_2SO_4 and concentrated. The crude residue is purified by flash chromatography (70 g SiO_2, DCM/MeOH/TEA 98.5:1.5:0.3 to 98:2:0.3) to yield a clear yellow film of **(3)** (2.0 g, 2.57 mmol, 74.8%). R_f = 0.32 (CH_2Cl_2/MeOH 19:1). ^{1}H NMR (500 MHz, $CDCl_3$) δ = 9.05 (s, 1H), 8.81 (s, 1H), 8.03-7.51 (m, 10H), 7.49-6.78 (m, 13H), 6.15 (m, 2H), 5.30 (s, 1H), 4.59 (m, 1H), 4.31 (m, 1H), 3.48 (m, 2H), 3.25 (m, 1H), 3.09 (m, 1H), 2.72 (m, 1H). HRMS calculated mass for $C_{45}H_{39}N_5O_6S$: 777.2621; found $[M+H]^+$: 778.2684.

N^4-Benzoyl-P(O)-(2-cyanoethyl)-5'-O-(4,4'-dimethoxytrityl)-2'-deoxycytidylyl-(3'→5')-N^6-benzoyl-3'-S-benzoyl-2',3'-dideoxyadenosine (4): First is to remove 5'-DMT from (3). To a solution of N^6-benzoyl-3'-S-benzoyl-2',3'-dideoxy-3'-thioadenosine (3) (2.05 g, 2.64

mmol, 1 eq.) in 50mL of DCM at 0 °C was added 25mL of 3% trifluoroacetic acid in DCM. The reaction was stirred for 7 min on ice and followed by TLC (DCM/MeOH 95:5). The color turns bright red upon addition of TFA. When complete, ca. 4 mL of MeOH was added followed by triethylamine. The reddish color turns back to light brown. The mixture is then extracted with 100mL DCM and 200mL brine, dried over $MgSO_4$, and concentrated. The crude residue is then purified by flash chromatography (70 g SiO_2, 98:2 DCM/MeOH) to yield 0.60 g of a light yellow solid (N^6-benzoyl-3'-S-benzoyl-2',3'-dideoxyadenosine, 1.26 mmol, 47.9%). ^1H NMR (500 MHz, $CDCl_3$) δ = 8.98 (s, 1H), 8.84 (s, 1H), 8.31 (m, 2H), 8.04-7.95 (m, 6H), 7.65-7.45 (m, 12H), 7.30-7.26 (m, 3H), 7.05 (m, 2H), 6.45-6.43 (m, 3H), 5.00-4.63 (m, 5H), 4.62-4.60 (m, 5H), 4.33 (m, 3H), 4.13-4.08 (m, 8H), 3.97-3.92 (m, 7H), 3.39-3.33 (m, 5H), 2.63-2.58 (m, 6H), 2.17-2.07 (m, 5H), 1.71-1.55 (m, 4H), 1.27-1.25 (m, 4H). HRMS calculated mass for $C_{24}H_{21}N_5O_4S$: 475.1314; found $[M+H]^+$: 476.1387.

In a dry 100 mL round-bottomed flask equipped with a magnetic stir bar, deprotected N^6-benzoyl-3'-S-benzoyl-2',3'-dideoxyadenosine (343 mg, 0.72 mmol, 1 eq.) was co-evaporated twice with 5 mL anhydrous pyridine and then dissolved in 8 mL anhydrous acetonitrile. 0.45M tetrazole in acetonitrile (8.0 mL, 3.6 mmol, 5 eq) was added to the flask and cooled to 0°C. N^4-Benzoyl-2'-deoxy-5'-O-DMT-cytidine 3'-CE phosphoramidite (660 mg, 0.79 mmol, 1.1 eq.) was dissolved in 8 mL of anhydrous acetonitrile added into mixture, and the reaction was stirred form 30 min at r.t. I_2 (220 mg, 0.87 mmol, 2.4 eq.) was dissolved in tetrahydrofuran/H_2O/pyridine (2:1:0.1) and added to the solution dropwisely. After stirring ca. 15 mins, the solution is poured into ca. 150 mL DCM and extracted with 100 mL $NaHSO_3$, the dark color of iodine disappeared. Collect organic phase and re-extract the aqueous phase with another 100 mL of DCM. The combined organic phase is washed with 200 mL brine, dried with Na_2SO_4 and concentrated. The crude residue is purified by

flash chromatography (120g SiO_2, DCM/MeOH/triethylamine 98:2:0.3) to give a yellow film. ^{1}H NMR (500 MHz, DMSO) δ = 9.23-9.21 (m, 2H), 8.91 (s, 1H), 8.49-8.41 (m, 2H), 8.09-8.08 (m, 2H), 8.00-7.91 (m, 4H), 7.69-7.51 (m, 10H), 7.49 (s, 1H), 7.41-7.14 (m, 13H), 6.90-6.89 (m, 2H), 6.59-6.55 (m, 4H), 6.37 (s, 1H), 5.36 (s, 1H), 5.25 (s, 1H), 4.69-6.68 (m, 2H), 4.52-4.42 (m, 5H), 4.31-4.14 (m, 7H), 3.83 (s, 1H), 3.54-3.49 (m, 4H), 3.36 (s, 1H), 2.86-2.68 (m, 9H), 2.41-2.39 (m, 2H), 2.32 (s, 1H), 2.13-2.03 (m, 3H), 1.61- 1.55 (m, 2H), 1.48-1.31 (m, 3H), 1.20 (s, 1H), 0.95-0.93 (m, 3H). HRMS calculated mass for $C_{64}H_{58}N_9O_{13}PS$: 1223.3612; found $[M+Na]^+$: 1246.3477.

N^4-Benzoyl-P(O)-(2-cyanoethyl)-2'-deoxycytidylyl-(3'→5')-N^6-benzoyl-3'-S-benzoyl-2',3'-dideoxyadenosine 5'-[bis(2-cyanoethyl)-phosphate] (5): To the solution of dinucleotide (4) (960 mg, 0.784 mmol, 1 eq.) in 20 mL of anhydrous dichloromethane at 0° C was added 10 mL of 3% TFA in DCM. The mixture was monitored by TLC (5:95 MeOH/DCM). After 30 min, the reaction was quenched with 5 mL of MeOH and poured into 75 mL DCM. The solution is washed with $NaHCO_3$ twice and brine once. The organic phase was dried by $MgSO_4$ and concentrated to give 860 mg crude residue. The crude product was purified by flash chromatography (60 g SiO_2, DCM/MeOH (from 97:3 to 93:7) to give 5'-deblocked dinucleotide (490 mg, 0.53 mmol, 67.8%). ^{1}H NMR (500 MHz, $CDCl_3$) δ = 9.37 (s, 1H), 9.18 (s, 1H), 8.86-8.78 (m, 2H), 8.53-7.92 (m, 10H), 7.64-7.52 (m, 3H), 7.08 (s, 1H), 6.58 (s, 1H), 6.26-6.17 (m, 2H), 5.37 (s, 1H), 5.16-3.87 (m, 14H), 3.55-3.19 (m, 3H), 2.83-2.68 (m, 2H), 2.49 (s, 1H), 1.28-1.12 (m, 2H). HRMS calculated mass for $C_{43}H_{40}N_9O_{11}PS$: 921.2306; found $[M+H]^+$: 922.2704.

In a dry 100 mL three-neck flask equipped with a magnetic stir bar, 5'-deblocked dinucleotide (490 mg, 0.53 mmol, 1 eq.) was co-evaporated twice with 5 mL anhydrous pyridine and then dissolved in 8 mL anhydrous acetonitrile. 0.45M tetrazole in acetonitrile (5.9 mL, 2.66

mmol, 5 eq.) was added to the flask and cooled to 0°C. Bis(2-cyanoethyl) diisopropylphosphoramidite (405 μL, 1.60 mmol, 3 eq.) was dissolved in 8 mL of anhydrous acetonitrile added into mixture, and the reaction was stirred form 30 min at r.t. I_2 (446 mg, 1.76 mmol, 3.3 eq.) was dissolved in tetrahydrofuran/H_2O/pyridine (2:1:0.1) and added to the solution dropwisely. After stirring ca. 15 mins, the solution is poured into ca. 100 mL DCM and extracted with 100 mL $NaHSO_3$, the dark color of iodine disappeared. Collect organic phase and re-extract the aqueous phase with another 100 mL of DCM. The combined organic phase is washed with 200 mL brine, dried with Na_2SO_4 and concentrated to yield pale yellow foam (580 mg, 0.52 mmol, 98.5%). The product is used for next step without further purification. ^{1}H NMR (500 MHz, $CDCl_3$) δ = 9.34 (s, 1H), 8.89-8.71 (m, 2H), 8.53-8.51 (m, 2H), 8.13-7.91 (m, 6H), 7.71-7.59 (m, 6H), 7.58-7.42 (m, 6H), 7.32-7.14 (m, 4H), 6.61 (s, 1H), 6.28-6.26 (m, 2H), 5.51 (s, 1H), 5.36 (s, 1H), 5.19-5.16 (m, 2H), 4.78-4.68 (m, 8H), 4.57-4.23 (m, 17H), 3.81 (s, 1H), 5.54-3.34 (m, 5H), 3.07 (s, 1H), 2.93-2.80 (m, 7H), 2.32-1.91 (m, 8H), 1.68 (s, 1H), 1.43-1.22 (m,7H). HRMS calculated mass for $C_{49}H_{47}N_{11}O_{14}P_2S$: 1107.2500; found $[M+H]^+$: 1108.2644.

2'-deoxycytidylyl-(3'→5')-3'-mercapto-2',3'-dideoxyadenosine 5'- dihydrogen phosphate Tris (triethyl-ammonium) Salt (6): In a pressure flask, dinucleotide **(5)** (50 mg, 0.045 mmol, 1 eq.) and *p*-MeO-toluenethiol (125 μL, 0.90 mmol, 20 eq. as scavenger) are dissolved in 2.5 mL of 1:9 dioxane/methanol. Concentrated ammonium hydroxide (2.5 mL, 29% ca.) was added to the reaction mixture and it was sealed and incubated at 55°C for 16 hr. The mixture is washed twice with 5 mL DCM to remove hydrophobic substance, and the aqueous phase is concentrated *in vacuuo* to give 39 mg crude residue. The crude was purified by reverse phase HPLC twice (Agilent Eclipse XDB-C18 column, 9.4 × 250 mm, first run: 0 to 50% linear gradient of acetonitrile in 50mM ammonium acetate over 50 min at flow rate 2 mL/min; second

run: 0 to 50% linear gradient of acetonitrile in 50mM triethylammonium acetate over 50 min at flow rate 2 mL/min) to yield product **(6)** (21.5 mg, 0.023 mmol, 50.7%). The product contains ca. 0.5 eq. of benzoic acid impurity, but this would not affect downstream reaction. ^{1}H NMR (600 MHz, D_2O) δ = 8.30 (s, 1H), 8.08 (s, 1H), 7.60 (d, 1H), 6.30 (d, 1H), 5.96 (q, 1H), 5.95 (m, 1H) 4.58-4.55 (m, 1H), 4.09-4.04 (m, 4H), 3.86-3.84 (m, 2H), 3.72 (q, 1H), 3.09 (q, 18H), 2.88-2.84 (m, 1H), 2.57-2.53 (m, 1H), 2.17-2.14 (m, 1H), 1.64-1.62 (m, 1H), 1.21-1.15 (t, 27H). HRMS calculated mass for $C_{19}H_{26}N_8O_{11}P_2S$: 636.0917; found $[M-H]^-$: 635.0748.

2'-deoxycytidylyl-(3'→5')-3'-S-malonyl-2',3'-dideoxyadenosine 5'- dihydrogen phosphate (7): To dinucleotide **(6)** (5mg, 7.86 μmol) in eppendorf tube is added 200μL of pH 7.1 sodium phosphate buffer (40mM). Malonyl-half-thioester (21 mg, 109 μmol, 13.9 eq) was dissolved in 50 μL acetonitrile and added to reaction. The mixture is titrated to pH 6 by 0.1N NaOH and vortex at r.t. overnight. The reaction progress is monitored by HPLC, and after completion, the mixture was quenched by 50 μL of 5% formic acid and purified by HPLC (0 to 25% linear gradient of acetonitrile in 0.1 % formic acid over 40 min at flow rate 10 mL/min) to yield (3.3 mg, 4.57 μmol, 58.1%). 1H NMR (600 MHz, D_2O) 8.48 (s, 1H), 8.38 (s, 1H), 8.22 (s, 1H), 7.83 (m, 2H), 6.40 (s, 1H), 6.09 (s, 1H), 5.98(s, 1H), 4.72 (m, 2H), 4.37 (s, 1H), 4.19 (s, 1H), 3.92 (m, 3H), 2.73 (m, 3H), 2.32 (s, 1H), 1.28(s, 1H). HRMS calculated for $C_{22}H_{28}N_8O_{14}P_2S$: 722.0920; found $[M+H]^+$: 723.0998.

Bis[2'-deoxycytidylyl-(3'→5')-3'-mercapto-2',3'-dideoxyadenosine 5'- dihydrogen phosphate)] disulfide (8): Dinucleotide **(5)** (200 mg, 0.180 mmol, 1 eq.) was deprotected following previous procedure. To the solution of crude dinucleotide thiol in 2 mL of tetrahydrofuran/ H_2O/pyridine (2:1:0.1) is added Iodine (91.2 mg, 0.180 mmol, 2 eq.) and stirred for 1 hr at r.t. The solution is washed with 2 mL DCM twice and cleaned up by HPLC (0 to 25%

linear gradient of acetonitrile in 0.1 % formic acid over 40 min at flow rate 10 mL/min). Yield: 51.7 mg (0.040 mmol, 45.2% over two steps) ^{1}H NMR (600 MHz, D_2O) δ = 8.41 (s, 2H) 8.22 (s, 2H) 7.90 (d, 2H), 6.37-6.35 (m, 2H) 6.07 (d, 2H), 5.91-5.89 (m, 2H), 4.67-4.62 (overlapped with reference signal, 2H expected), 4.27-4.25 (m, 2H), 4.16-4.10 (m, 4H), 4.06-4.03 (m, 2H), 3.91-3.83 (m, 6H), 2.98-2.97 (m, 2H), 2.80-2.76 (m, 2H), 2.37-2.33 (m, 2H), 1.94-1.90 (m, 2H). HRMS calculated mass for $C_{38}H_{50}N_{16}O_{22}P_4S_2$: 1270.1677; found $[M+H]^+$: 1271.1721.

2'-deoxycytidylyl-(3'→5')-3'-S-*p*-hydroxyphenacyl-2',3'-dideoxyadenosine 5'-dihydrogen phosphate (9): Dinucleotide **(5)** (200 mg, 0.180 mmol, 1 eq.) was deprotected following previous procedure. To the solution of crude dinucleotide thiol in 2 mL of CH_3CN/phosphate buffer pH 7.1 (1:1) is added 4-hydrophenacyl bromide (31.0 mg, 0.144 mmol, 0.8 eq). The solution is stirred at r.t. overnight, and purified by HPLC (0 to 25% linear gradient of acetonitrile in 0.1 % formic acid over 40 min at flow rate 10 mL/min). Yield: ca. 84.5 mg (0.110 mmol, 60.9% over two steps) ^{1}H NMR (600 MHz, D_2O) δ = 8.27 (s, 1H), 8.23 (s, 1H), 8.05 (s, 1H), 7.76-7.73 (m, 2H), 6.72-6.71 (), 6.20-6.18 (m, 1H), 5.94 (d, 1H), 5.81-5.79 (dd, 1H), 4.14-4.11 (m, 2H), 4.09-4.06 (m, 2H), 4.01-3.99 (m, 1H), 3.96-3.94 (d, 1H), 3.85 (m, 2H), 3.52-3.48 (m, 1H), 2.69-2.67 (m, 1H), 2.49-2.47 (m, 1H), 2.25-2.23 (m, 1H), 1.82-1.77 (m, 1H). HRMS calculated mass for $C_{27}H_{32}N_8O_{13}P_2S$: 770.1285; found $[M+H]^+$: 771.1344.

Photolysis of *p*HP-protected thiol:

In clear eppendorf tube, 50 µL dinucleotide-thiol-*p*HP **(9)**, 20 µL $NaBO_4$ buffer (0.5 M, pH 8.0), 10 µL 0.5M DTT, and 20 µL Milli-Q water were combined and illuminated under 350 nm UV light (21000µW/cm2, 3 min) on ice 0°C. The product is then analyzed by LCMS to confirm dinucleotide thiol **(6)** is recovered.

References

Ellman, J., Mendel, D., Anthony-Cahill, S., Noren, C.J., and Schultz, P.G. (1991). [15] Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. Methods Enzymol. *202*, 301–336.

Eun, H.M. (1996). Enzymology primer for recombinant DNA technology (Academic Pr).

Ferrer, J.L., Jez, J.M., Bowman, M.E., Dixon, R.A., and Noel, J.P. (1999). Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. Nat. Struct. Mol. Biol. *6*, 775–784.

Gindulyte, A., Bashan, A., Agmon, I., Massa, L., Yonath, A., and Karle, J. (2006). The transition state for formation of the peptide bond in the ribosome. Proc. Natl. Acad. Sci. U. S. A. *103*, 13327–13332.

Goess, B.C., Hannoush, R.N., Chan, L.K., Kirchhausen, T., and Shair, M.D. (2006). Synthesis of a 10,000-membered library of molecules resembling carpanone and discovery of vesicular traffic inhibitors. J. Am. Chem. Soc. *128*, 5391–5403.

McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., and Ashley, G. (1999). Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel “unnatural” natural products. Proc. Natl. Acad. Sci. U. S. A. *96*, 1846–1851.

Menzella, H.G., Reid, R., Carney, J.R., Chandran, S.S., Reisinger, S.J., Patel, K.G., Hopwood, D. a, and Santi, D. V (2005). Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. Nat. Biotechnol. *23*, 1171–1176.

Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C., and Schultz, P.G. (1989). A general method for site-specific incorporation of unnatural amino acids into proteins. Science *244*, 182–188.

Porcher, S., Meyyappan, M., and Pitsch, S. (2005). Spontaneous Aminoacylation of a RNA Sequence Containing a 3′-Terminal 2′-Thioadenosine. Helv. Chim. Acta *88*, 2897–2909.

Voorhees, R.M., Weixlbaumer, A., Loakes, D., Kelley, A.C., and Ramakrishnan, V. (2009). Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. Nat. Struct. Mol. Biol. *16*, 528–533.

Weissman, K.J., and Leadlay, P.F. (2005). Combinatorial biosynthesis of reduced polyketides. Nat. Rev. Microbiol. *3*, 925–936.

Xue, Q., Ashley, G., Hutchinson, C.R., and Santi, D. V (1999). A multiplasmid approach to preparing large libraries of polyketides. Proc. Natl. Acad. Sci. U. S. A. *96*, 11740–11745.

Appendix

Appendix A: Python codes

Python code for generating random insertion and deletion oligo sequences

```
#!/usr/bin/env python
from Bio import SeqIO
from Bio.Seq import Seq
from Bio.SeqRecord import SeqRecord
from Bio import SeqFeature
from Bio.Seq import MutableSeq
from Bio.Alphabet import generic_dna

def RID(wt,ins, det, LibSize):
    #ins = rate of insertion(&)
    #det = rate of deletion(%)
    #LibSize = Number of oligos generated
    import random
    inscnt = 0
    delcnt = 0
    mutcnt = 0
    j = 0
    while j < LibSize:
        target = wt.tomutable()
        ghost = wt.tomutable()
        i = 0
        k = 0
        while i < len(target):
            R = random.random()
            if R < ins*det:                         # deletion + insertion = randomize
                target[i:(i+1)] = random.choice('atgc')
                ghost[k:k+1] = "N"
                mutcnt = mutcnt + 1
                k = k + 1
                i = i + 1
            elif ins*det <= R < ins + ins*det:      # insertion
                target[i:(i+1)] = target[i] + random.choice('atgc')
                ghost[k:k+1] = "&"
                inscnt = inscnt + 1
                k = k + 1
                i = i + 2
            elif (1-det) <= R:                      # deletion
                target[i:(i+1)] = ""
                ghost[k:k+1] = "%"
                delcnt = delcnt + 1
                k = k + 1
                i = i
            else:
                ghost[k:k+1] = "-"
                k = k + 1
                i = i + 1
        #print wt, ghost, target
        j = j + 1
        return [target, ghost]
    #print "total insertion = ",inscnt
    #print "total deletion = ",delcnt
    #print "total randomization", mutcnt
frag = [1,2,3,4,5,6,7,8]
libsize = [4000, 10000, 4000, 4500, 23000, 4500, 20000, 20000]
RIDfrag = 1
```

```
lib = libsize[RIDfrag-1]

handle = open("C:\\Users\\Poyi\\Documents\\170mer-oligo-"+ str(RIDfrag) +".gbk", "rU")
for record in SeqIO.parse(handle, "genbank") :
    #print record.format('fasta')
    #print record.features
    id = record.id
handle.close()
myreport = []
k = 0
while k < lib:
    mutrecord = record
    ghorecord = record
    framelist = []
    for x in range(len(record.seq)):
        framelist.append(0)
    #print framelist
    mutframe = 0
    ghoframe = 0
    for feat in record.features:
        if feat.type == 'misc_feature':                 #pick the mutation features
            start = feat.location.start
            end = feat.location.end
            featseq = record[start:end]                 #extract feature seq
            #print featseq
            #print featseq.seq
            #print start, end
            mut = RID(featseq.seq, 0.03, 0.05, 1)       #replace feature with RID seq
            #print mut[0]
            mutrecord.id = id + '-mut' + str(k+1)
            z = []                  #amend the framelist right AFTER the end of feature
            for y in framelist[end:]:
                z = z + [y + (len(mut[0]) - len(featseq.seq))]
            framelist[end:] = z
            #print framelist
            mutrecord = mutrecord[:start + framelist[start]] + mut[0] + mutrecord[end
+ framelist[start]:]  #can add start-end frame check later.
            #print mutrecord
            ghorecord = ghorecord[:start + ghoframe] + mut[1] + ghorecord[end +
ghoframe:]
            mutframe = mutframe + (len(mut[0]) - len(featseq.seq))
            ghoframe = ghoframe + (len(mut[1]) - len(featseq.seq))
    #print ghorecord.seq
    if len(mutrecord.seq) > 170:
        k = k
    else:
        myreport.append(mutrecord)
        print mutrecord.format('fasta')
        k = k + 1
SeqIO.write(myreport, "170mer-"+str(RIDfrag)+"-mut.fasta", 'fasta')
```

Appendix B: Sequences

Sequences of templates used in ribosome mutant selection

$DNA_{1/12}$:

GGCGTAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATggattacaagga
tgacgacgataagatcAAAagtctgattTAGgtagatcgcgttatcTGCatggaaaacgccatgTAGtggaacctgcctgccgat
ctcgcctggtttaaaTAGaacaccttaaataaacccgtgattatgTGCcgcTAGacctgggaatcaatcCATcgtccgttgccaG
CGTAGaaaaatATTATCCTCAGCAGTCAACCGCATACGTAGGATCGCGTAACGTGGGTG
AAGTCGGTGGATGAATAGATCGCGGCGTGTCATGACGTACCAGAAATCATGTAGAT
TGCGGCGCATCGCGTTTATGAACAGTTCTTGTAGAAGGCGCAAAAACTGTATCTGAC
GCATATCGACTAGGAAGTGGAAGCGGACACCCATTTCCCGGATTACTAGCCGGATG
ACTGGGAATCGGTATTCAGCGAATTCTAGGATGCTGATGCGCAGAACTCT

$DNA_{1/6}$:

GGCGTAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATggattacaagga
tgacgacgataagatcAAAagtctgattTAGgtagatcgcgttatcTAGatggaaaacgccatgTAGtggaacctgcctgccTA
Gctcgcctggtttaaa TAGaacaccttaaataaTAGgtgattatgTGCcgcTAGacctgggaatcaatcTAGcgtccgttgcc
aGCGTAGaaaaatATTATCCTCTAGAGTCAACCGCATACGTAGGATCGCGTAACGTGGT
AGAAGTCGGTGGATGAATAGATCGCGGCGTGTCATTAGGTACCAGAAATCATGTAG
ATTGCGGCGCATCGCTAGTATGAACAGTTCTTGTAGAAGGCGCAAAAACTGTAGCTG
ACGCATATCGACTAGGAAGTGGAAGCGGACTAGCATTTCCCGGATTACTAGCCGGA
TGACTGGGAATAGGTATTCAGCGAATTCTAGGATGCTGATGCGCAGTAGTCT

$DNA_{3/6}$:

GGCGTAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATggattacaagga
tgacgacgataagatcAAAagtctgattTAGgtagatTAGgttatcTAGatggaaTAGgccatgTAGtggaacTAGcctg
ccTAGctcgccTAGtttaaaTAGaacaccTAGaataaaTAGgtgattTAGTGCcgcTAGacctggTAGtcaatcT
AGcgtccgTAGccaGCGTAGaaaaatTAGATCCTCTAGAGTCAATAGCATACGTAGGATCGC
TAGACGTGGTAGAAGTCGTAGGATGAATAGATCGCGTAGTGTCATTAGGTACCATA
GATCATGTAGATTGCGTAGCATCGCTAGTATGAATAGTTCTTGTAGAAGGCGTAGAA
ACTGTAGCTGACGTAGATCGACTAGGAAGTGTAGGCGGACTAGCATTTCTAGGATTA
CTAGCCGGATTAGTGGGAATAGGTATTCTAGGAATTCTAGGATGCTTAGGCGCAGTA
GTCT

$DNA_{2/6}$:

GGCGTAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATggattacaagga
tgacgacgataagatcAAAagtctgTAGTAGgtagatcgcgttTAGTAGatggaaaacgccTAGTAGtggaacctgcct
TAGTAGctcgcctggtttTAGTAGaacaccttaaatTAGTAGgtgattatgTGCTAGTAGacctgggaatcaTAG
TAGcgtccgttgccaTAGTAGaaaaatATTATCTAGTAGAGTCAACCGCATTAGTAGGATCGCG
TAACGTAGTAGAAGTCGGTGGATTAGTAGATCGCGGCGTGTTAGTAGGTACCAGAA
ATCTAGTAGATTGCGGCGCATTAGTAGTATGAACAGTTCTAGTAGAAGGCGCAAAA
ATAGTAGCTGACGCATATCTAGTAGGAAGTGGAAGCGTAGTAGCATTTCCCGGATTA
GTAGCCGGATGACTGGTAGTAGGTATTCAGCGATGCTGATGCGCAGTAGTCT

$DNA_{3/12}$:

GGCGTAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATggattacaagga
tgacgacgataagatcAAAagtTAGTAGTAGgtagatcgcgttatcTGCatggaaaacTAGTAGTAGtggaacctgc
ctgccgatctcgcctggTAGTAGTAGaacaccttaaataaacccgtgattatgTAGTAGTAGgcctgggaatcaatcCAT
cgtccgttgTAGTAGTAGaaaaatATTATCCTCAGCAGTCAACCGTAGTAGTAGGATCGCGTA

ACGTGGGTGAAGTCGGTGTAGTAGTAGATCGCGGCGTGTCATGACGTACCAGAATA
GTAGTAGcTTGCGGCGCATCGCGTTTATGAACAGTAGTAGTAGAAGGCGCAAAAACT
GTATCTGACGCATTAGTAGTAGGAAGTGGAAGCGGACACCCATTTCCCGTAGTAGTA
GCCGGATGACTGGGAATCGGTATTCAGCGATGCTGATGCGCAGAACTCT

Appendix C: Lists of Mutations in Selected Mutants

Mutations in mutant G2-G5, NP1-NP5 (whether it is programmed by chip oligos)

RID-G1

A592T (No)
T596A (No)
A599G (No)
A602G (No)
G605C (No)
C623G (No)
A626T (No)
G629C (No)
C634G (No)
A637G (No)
C645T (No)
T646A (No)
T653C (No)
A654T (No)
T658C (No)
A661T (No)
T665A (No)
C680G (No)
T683C (No)
G690T (No)
A743C (Yes)
T744C (Yes)
G745T (Yes)
A753G (Yes)
T766C (No)
C772A (No)
G774C (No)
G777A (Yes)
C787T (Yes)
A794G (No)
C796T (No)
G797C (No)
C816T (No)
G841A (No)
A844T (No)
T846- (No)
T850C (No)
C851T (No)
T852G (No)
C853T (No)
C854A (No)
A877T (No)
C901A (No)
G914C (No)
G923T (No)
G924A (No)
A925C (No)
G926A (No)
A928G (No)
T931C (No)
T932C (No)
-932G (No)
T934G (No)
C937T (No)
G974A (No)
C994T (No)
C1013T (No)
C1044T (No)
C1045T (No)
C1053T (No)
T1058G (No)
T1078C (No)
A1080C (No)
G1106A (No)
-1141C (No)
G1149A (No)
G1160A (No)
C1164- (No)
A1165- (No)
C1167T (No)
G1168A (No)
A1169T (No)
G1171A (No)
C1172T (No)
T1173C (No)
A1175T (No)
T1176C (No)
C1178G (No)
G1179A (No)
T1181G (No)
G1182A (No)
T1184G (No)
G1185C (No)
G1191A (No)
C1211T (No)
G1220A (No)
C1229T (No)
A1284- (No)
T1513C (No)
G1516A (No)
G1537T (No)
C1541T (No)
G1581C (No)
G1723A (No)
A1784T (Yes)
G1807A (No)
G1869T (No)
C1870A (No)
A1872C (No)
T1915G (No)
T1979C (No)
C1986G (No)
A1987G (No)
G1988C (No)
C2021A (No)
T2132A (No)
T2203G (No)
A2211G (No)
total number of mutations = 108

RID-G2

A592T (No)
T596A (No)
A599G (No)
A602G (No)
G605C (No)
C623G (No)
A626T (No)
G629C (No)
C634G (No)
C645T (No)
T646A (No)
T653C (No)
A654T (No)
T658C (No)
A661T (No)
T665A (No)
C680G (No)
T683C (No)
G690T (No)
A743C (Yes)
T744C (Yes)
G745T (Yes)
A753G (Yes)
T766C (No)
C772A (No)
G774C (No)
G777A (Yes)
C787T (Yes)
A794G (No)
C796T (No)
G797C (No)
C816T (No)
G841A (No)
A844T (No)
T846- (No)
T850C (No)
C851T (No)
T852G (No)
C853T (No)
C854A (No)
A877T (No)
C901A (No)
G914C (No)
G923T (No)
G924A (No)
A925C (No)
G926A (No)
A928G (No)
T931C (No)
T932C (No)
-932G (No)
T934G (No)
C937T (No)
G974A (No)
C994T (No)
C1013T (No)
C1044T (No)
C1045T (No)
C1053T (No)
T1058G (No)
T1078C (No)
A1080C (No)
G1106A (No)
-1141C (No)
G1149A (No)
G1160A (No)
C1164- (No)
A1165- (No)
C1167T (No)
G1168A (No)
A1169T (No)
G1171A (No)
C1172T (No)
T1173C (No)
A1175T (No)
T1176C (No)
C1178G (No)
G1179A (No)
T1181G (No)
G1182A (No)
T1184G (No)
G1185C (No)
G1191A (No)
C1211T (No)
T1219A (No)
C1221T (No)
T1224A (No)
G1227A (No)
A1230T (No)
G1245A (No)
A1264G (No)
A1276T (No)
C1278G (No)
G1280T (No)
G1288A (No)
C1290A (No)
G1292C (No)
T1294A (No)
G1300A (No)
G1311T (No)
T1318C (No)
C1319G (No)
G1332A (No)
G1333C (No)
G1341T (No)
A1347G (No)
C1348T (No)
C1363T (No)
G1368A (No)
G1382A (No)
G1407C (No)
G1408T (No)
T1409G (No)
C1414T (No)
A1420T (No)
G1423A (No)
T1440G (No)
T1442C (No)
T1443C (No)
G1444A (No)
C1447T (No)
G1448T (No)
A1453T (No)
C1454T (No)
C1463A (No)
G1464A (No)
C1472G (No)
G1473T (No)
T1474G (No)
T1484A (No)
T1485A (No)
T1487C (No)

C1488T (No)
C1489T (No)
C1493T (No)
A1502G (No)
A1504T (No)
A1505T (No)
T1506C (No)
C1507T (No)
T1513C (No)
G1517A (No)
C1518G (No)
G1519C (No)
G1530T (No)
-1532T (No)
-1532C (No)
-1532T (No)
C1533T (No)
A1535T (No)
C1536A (No)
G1538A (No)
C1541A (No)
T1542C (No)
C1547T (No)
A1548G (No)
A1549G (No)
C1550T (No)
A1551T (No)
A1552G (No)
C1558A (No)
G1567A (No)
C1575T (No)
A1583T (No)
A1593C (No)
T1594A (No)
C1595G (No)
A1596G (No)
T1599C (No)
G1723A (No)
T1725C (No)
C1726G (No)
C1727A (No)
C1730T (No)
G1733T (No)
G1734C (No)
A1735G (No)
G1792C (No)
T1915G (No)
total number of mutations = 179

A654T (No)
T658C (No)
A661T (No)
T665A (No)
C680G (No)
T683C (No)
G690T (No)
A743C (Yes)
T744C (Yes)
G745T (Yes)
A753G (Yes)
T766C (No)
C772A (No)
G774C (No)
G777A (Yes)
C787T (Yes)
A794G (No)
C796T (No)
G797C (No)
C816T (No)
C838A (No)
G841A (No)
A844T (No)
T846- (No)
T850C (No)
C851T (No)
T852G (No)
C853T (No)
C854A (No)
A877T (No)
C901A (No)
C908T (No)
G914C (No)
G923T (No)
G924A (No)
A925C (No)
G926A (No)
A928G (No)
T931C (No)
T932C (No)
-932G (No)
T934G (No)
C937T (No)
G974A (No)
C994T (No)
C1013T (No)
C1044T (No)
C1045T (No)
C1053T (No)
T1058G (No)
T1078C (No)
A1080C (No)
G1106A (No)
-1141C (No)
G1149A (No)
G1160A (No)
C1164- (No)
A1165- (No)
C1167T (No)
G1168A (No)
A1169T (No)
G1171A (No)
C1172T (No)

T1173C (No)
A1175T (No)
T1176C (No)
C1178G (No)
G1179A (No)
T1181G (No)
G1182A (No)
T1184G (No)
G1185C (No)
G1191A (No)
C1211T (No)
T1219A (No)
C1221T (No)
T1224A (No)
G1227A (No)
A1230T (No)
A1264G (No)
A1276T (No)
C1278G (No)
G1280T (No)
G1288A (No)
C1290A (No)
G1292C (No)
T1294A (No)
G1300A (No)
G1311T (No)
T1318C (No)
C1319G (No)
G1332A (No)
G1333C (No)
G1341T (No)
A1347G (No)
C1348T (No)
C1363T (No)
G1368A (No)
G1382A (No)
G1407C (No)
G1408T (No)
T1409G (No)
C1414T (No)
A1420T (No)
G1423A (No)
T1440G (No)
T1442C (No)
T1443C (No)
G1444A (No)
C1447T (No)
G1448T (No)
A1453T (No)
C1454T (No)
C1463A (No)
G1464A (No)
C1472G (No)
G1473T (No)
T1474G (No)
T1484A (No)
T1485A (No)
T1487C (No)
C1488T (No)
C1489T (No)
C1493T (No)
A1502G (No)
A1504T (No)

A1505T (No)
T1506C (No)
C1507T (No)
T1513C (No)
G1517A (No)
C1518G (No)
G1519C (No)
G1530T (No)
-1532T (No)
-1532C (No)
-1532T (No)
C1533T (No)
A1535T (No)
C1536A (No)
G1538A (No)
C1541A (No)
T1542C (No)
C1547T (No)
A1548G (No)
A1549G (No)
C1550T (No)
A1551T (No)
A1552G (No)
C1558A (No)
G1567A (No)
C1575T (No)
A1583T (No)
A1593C (No)
T1594A (No)
C1595G (No)
A1596G (No)
T1599C (No)
C1625T (No)
A1626G (No)
T1683C (No)
G1684A (No)
C1704T (No)
A1705G (No)
T1709C (No)
A1711- (No)
T1712- (No)
A1713G (No)
T1716A (No)
A1717G (No)
G1723A (No)
T1725- (No)
C1726G (No)
C1727A (No)
C1730T (No)
G1733T (No)
G1734C (No)
A1735C (No)
T1736G (No)
G1737T (No)
G1738A (No)
G1743C (No)
A1745T (No)
A1746G (No)
T1747C (No)
A1749G (No)
C1764G (No)
T1765C (No)
G1766C (No)

A1772G (No)
G1797C (No)
C1822G (No)
G1869T (No)
C1870A (No)
A1872C (No)
T1915G (No)
T1979C (No)
C1986G (No)
A1987G (No)
G1988C (No)
C2021A (No)
C2043T (No)
A2051T (No)
G2087T (No)
A2088G (No)
A2094G (No)
A2097T (No)
G2102A (No)
C2103T (No)
C2104T (No)
T2106G (No)
G2107C (No)
A2108T (No)
T2137C (No)
C2150T (No)
T2151C (No)
A2154G (No)
G2162A (No)
T2181G (No)
T2182G (No)
A2183C (No)
T2185A (No)
G2186C (No)
T2192G (No)
G2201T (No)
T2202C (No)
T2203A (No)
A2205G (No)
C2206T (No)
A2211T (No)
G2218A (No)
T2220C (No)
C2222A (No)
C2232A (No)
C2283T (No)
T2292G (No)
G2293C (No)
A2297C (No)
T2299G (No)
C2300A (No)
T2302C (No)
C2310A (No)
A2314G (No)
G2316T (No)
A2317C (No)
G2319T (No)
T2320A (No)
T2321G (No)
G2325A (No)
C2326T (No)
T2329A (No)
T2334A (No)

RID-G3

A592T (No)
T596A (No)
A599G (No)
A602G (No)
G605C (No)
C623G (No)
A626T (No)
G629C (No)
C634G (No)
C645T (No)
T646A (No)
T653C (No)

C2339G (No)
A2340C (No)
C2354A (No)
G2355C (No)
T2356A (No)
G2360A (No)
G2361C (No)
C2362G (No)
G2363T (No)
T2533C (No)
A2534C (No)
total number of mutations = 275

RID-G4

A592T (No)
T596A (No)
A599G (No)
A602G (No)
G605C (No)
C623G (No)
A626T (No)
G629C (No)
C634G (No)
C645T (No)
T646A (No)
T653C (No)
A654T (No)
T658C (No)
A661T (No)
T665A (No)
C680G (No)
T683C (No)
G690T (No)
A743C (Yes)
T744C (Yes)
G745T (Yes)
A753G (Yes)
T766C (No)
C772A (No)
G774C (No)
G777A (Yes)
C787T (Yes)
A794G (No)
C796T (No)
G797C (No)
C816T (No)
G841A (No)
A844T (No)
T846- (No)
T850C (No)
C851T (No)
T852G (No)
C853T (No)
C854A (No)
A877T (No)
C901A (No)
G914C (No)
G923T (No)
G924A (No)
A925C (No)
G926A (No)
A928G (No)
T931C (No)
T932C (No)
-932G (No)
T934G (No)
C937T (No)
G974A (No)
C994T (No)
C1013T (No)
C1044T (No)
C1045T (No)
C1053T (No)
T1058G (No)
T1078C (No)
A1080C (No)
G1106A (No)
-1141C (No)
G1149A (No)
G1160A (No)
C1164- (No)
A1165- (No)
C1167T (No)
G1168A (No)
A1169T (No)
G1171A (No)
C1172T (No)
T1173C (No)
A1175T (No)
T1176C (No)
C1178G (No)
G1179A (No)
T1181G (No)
G1182A (No)
T1184G (No)
G1185C (No)
G1191A (No)
C1211T (No)
T1219A (No)
C1221T (No)
T1224A (No)
G1227A (No)
A1230T (No)
A1264G (No)
A1276T (No)
C1278G (No)
G1280T (No)
G1288A (No)
C1290A (No)
G1292C (No)
T1294A (No)
G1300A (No)
G1311T (No)
T1318C (No)
C1319G (No)
G1332A (No)
G1333C (No)
G1341T (No)
A1347G (No)
C1348T (No)
C1363T (No)
G1368A (No)
G1382A (No)
G1407C (No)
G1408T (No)
T1409G (No)
C1414T (No)
A1420T (No)
G1423A (No)
T1440G (No)
T1442C (No)
T1443C (No)
G1444A (No)
C1447T (No)
G1448T (No)
A1453T (No)
C1454T (No)
C1463A (No)
G1464A (No)
C1472G (No)
G1473T (No)
T1474G (No)
T1484A (No)
T1485A (No)
T1487C (No)
C1488T (No)
C1489T (No)
C1493T (No)
A1502G (No)
A1504T (No)
A1505T (No)
T1506C (No)
C1507T (No)
T1513C (No)
G1517A (No)
C1518G (No)
G1519C (No)
G1530T (No)
-1532T (No)
-1532C (No)
-1532T (No)
C1533T (No)
A1535T (No)
C1536A (No)
G1538A (No)
C1541A (No)
T1542C (No)
C1547T (No)
A1548G (No)
A1549G (No)
C1550T (No)
A1551T (No)
A1552G (No)
C1558A (No)
G1567A (No)
C1575T (No)
A1583T (No)
A1593C (No)
T1594A (No)
C1595G (No)
A1596G (No)
T1599C (No)
C1625T (No)
A1626G (No)
T1683C (No)
G1684A (No)
C1704T (No)
A1705G (No)
T1709C (No)
A1711- (No)
T1712- (No)
A1713G (No)
T1716A (No)
A1717G (No)
G1723A (No)
T1725- (No)
C1726G (No)
C1727A (No)
C1730T (No)
G1733T (No)
G1734C (No)
A1735C (No)
T1736G (No)
G1737T (No)
G1738A (No)
G1743C (No)
A1745T (No)
A1746G (No)
T1747C (No)
A1749G (No)
C1764G (No)
T1765C (No)
G1766C (No)
A1772G (No)
G1797C (No)
C1822G (No)
G1869T (No)
C1870A (No)
A1872C (No)
T1915G (No)
T1979C (No)
C1986G (No)
A1987G (No)
G1988C (No)
C2043T (No)
G2140A (No)
T2203G (No)
A2211G (No)
T2629C (No)
total number of mutations = 215

RID-G5

A592T (No)
T596A (No)
A599G (No)
A602G (No)
G605C (No)
C623G (No)
A626T (No)
G629C (No)
C634G (No)
C645T (No)
T646A (No)
T653C (No)
A654T (No)
T658C (No)
A661T (No)
T665A (No)
C680G (No)
T683C (No)
G690T (No)
A743C (Yes)
T744C (Yes)
G745T (Yes)
A753G (Yes)
T766C (No)
C772A (No)
G774C (No)
G777A (Yes)
C787T (Yes)
A794G (No)
C796T (No)
G797C (No)
C816T (No)
G841A (No)
A844T (No)
T846- (No)
T850C (No)
C851T (No)
T852G (No)
C853T (No)
C854A (No)
A877T (No)
C901A (No)
G914C (No)
G923T (No)
G924A (No)
A925C (No)
G926A (No)
A928G (No)
T931C (No)
T932C (No)
-932G (No)
T934G (No)
C937T (No)
G974A (No)
C994T (No)
C1013T (No)
C1044T (No)
C1045T (No)
C1053T (No)
T1058G (No)
T1078C (No)
A1080C (No)
G1106A (No)
-1141C (No)
G1149A (No)
G1160A (No)
C1164- (No)
A1165- (No)
C1167T (No)
G1168A (No)
A1169T (No)
G1171A (No)
C1172T (No)
T1173C (No)
A1175T (No)
T1176C (No)
C1178G (No)
G1179A (No)
T1181G (No)
G1182A (No)
T1184G (No)

G1185C (No)
G1191A (No)
C1211T (No)
T1219A (No)
C1221T (No)
T1224A (No)
G1227A (No)
A1230T (No)
A1264G (No)
A1276T (No)
C1278G (No)
G1280T (No)
G1288A (No)
C1290A (No)
G1292C (No)
T1294A (No)
G1300A (No)
G1311T (No)
T1318C (No)
C1319G (No)
G1332A (No)
G1333C (No)
G1341T (No)
A1347G (No)
C1348T (No)
C1363T (No)
G1368A (No)
G1382A (No)
G1407C (No)
G1408T (No)
T1409G (No)
C1414T (No)
A1420T (No)
G1423A (No)
T1440G (No)
T1442C (No)
T1443C (No)
G1444A (No)
C1447T (No)
G1448T (No)
A1453T (No)
C1454T (No)
C1463A (No)
G1464A (No)
C1472G (No)
G1473T (No)
T1474G (No)
T1484A (No)
T1485A (No)
T1487C (No)
C1488T (No)
C1489T (No)
C1493T (No)
A1502G (No)
A1504T (No)
A1505T (No)
T1506C (No)
C1507T (No)
T1513C (No)
G1517A (No)
C1518G (No)
G1519C (No)
G1530T (No)
-1532T (No)
-1532C (No)
-1532T (No)
C1533T (No)
A1535T (No)
C1536A (No)
G1538A (No)
C1541A (No)
T1542C (No)
C1547T (No)
A1548G (No)
A1549G (No)
C1550T (No)
A1551T (No)
A1552G (No)
C1558A (No)
G1567A (No)
C1575T (No)
A1583T (No)
A1593C (No)
T1594A (No)
C1595G (No)
A1596G (No)
T1599C (No)
C1625T (No)
A1626G (No)
T1683C (No)
G1684A (No)
C1704T (No)
A1705G (No)
T1709C (No)
A1711- (No)
T1712- (No)
A1713G (No)
T1716A (No)
A1717G (No)
G1723A (No)
T1725- (No)
C1726G (No)
C1727A (No)
C1730T (No)
G1733T (No)
G1734C (No)
A1735C (No)
T1736G (No)
G1737T (No)
G1738A (No)
G1743C (No)
A1745T (No)
A1746G (No)
T1747C (No)
A1749G (No)
C1764G (No)
T1765C (No)
G1766C (No)
A1772G (No)
G1797C (No)
C1822G (No)
G1869T (No)
C1870A (No)
A1872C (No)
T1915A (No)
T1979C (No)
C1986G (No)
A1987G (No)
G1988C (No)
C2021A (No)
T2203G (No)
A2211G (No)
total number of mutations = 213

RID-NP1

G745C (Yes)
A1134T (No)
C1164T (No)
C1178T (No)
C1196T (No)
G1831A (No)
T1915A (No)
G1989A (No)
G2487A (No)
total number of mutations = 9

RID-NP2

G745T (Yes)
A1230G (No)
C1357T (No)
T1915C (No)
T2132A (No)
T2203G (No)
A2211G (No)
C2374T (No)
total number of mutations = 8

RID-NP3

G745T (Yes)
C1150T (No)
A1269T (No)
G1723A (No)
T1865C (No)
T1915C (No)
T2203G (No)
A2211G (No)
G2444- (Yes)
G2445- (Yes)
C2575T (Yes)
total number of mutations = 11

RID-NP4

G745T (Yes)
C915T (No)
G1171A (No)
T1173- (No)
C1178T (No)
C1211T (No)
G1220A (No)
C1229T (No)
T1484G (No)
T1513C (No)
C1541T (No)
G1723A (No)
T1915G (No)
T2132A (No)
T2166C (No)
G2623T (No)
total number of mutations = 16

RID-NP5

A626G (No)
T744- (Yes)
G745T (Yes)
C957T (Yes)
G1171A (No)
T1173- (No)
A1175G (No)
C1178T (No)
C1211T (No)
G1220A (No)
C1229T (No)
G1723A (No)
C1730T (No)
G1869T (No)
C1870A (No)
A1872C (No)
-1891C (No)
T1915G (No)
C1974T (No)
T1979C (No)
C1986G (No)
A1987G (No)
G1988C (No)
T2493C (Yes)
total number of mutations = 24